

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:37:40 : Search time 276.658 Seconds
(without alignments) 9873.841 Million cell updates/sec

Title: US-09-831-426C-4

Perfect score: 1213

Sequence: 1 gtgcgcgcgcgcgcgcgaag.....cagtactactccttgctta 1213

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1213	100.0	1213	21	AAA15407
2	1211.4	99.9	1273	21	AAA15405
3	1206.2	99.4	1349	22	AAH34586
4	1171.6	96.6	1381	24	ABR83650
5	1170	96.5	1399	17	AA114038
6	1167	96.2	1269	17	AA114037
7	495	40.8	690	24	ABO54923
8	418.4	34.5	439	24	ABK44775
9	416.8	34.4	439	24	ABR45138

10	374	30.8	537	24	ABO59499	Human colon cancer
11	301.4	24.8	466	23	ABV49231	Human prostate exp
12	300.4	24.8	425	23	ABV19459	Human prostate exp
13	246	20.3	1401	22	AAH22129	Human eukaryotic i
14	236	19.5	1889	23	AA573554	DNA encoding novel
15	234	19.3	234	24	ABR45059	CDNA encoding colo
16	193	15.9	291	16	AA19568	CDNA encoding colo
17	177	14.6	272	24	AA159173	Human gene signatu
18	152.4	12.6	660	22	AAH23370	Activated T-cell d
19	152.4	12.6	947	22	AAH15428	Nucleotide sequenc
20	152.4	12.6	947	22	AAAB9338	DNA encoding zinc
21	152.4	12.6	947	22	AAAC91046	Plasmid pTR11A2if
22	152.4	12.6	995	22	AA511447	Vector pTR11A/2if
23	152.4	12.6	995	22	AAAB9337	DNA encoding zinc
24	152.4	12.6	995	22	AAAC91045	Plasmid pTR11A2if
25	101.6	8.4	183	24	ABN24598	Vector pTR11A/2if
26	97	8.0	919	22	AAH48496	Human ORKX poly/nuc
27	91.2	7.5	3302	24	AB169125	Japanese medaka te
28	81	6.7	368	24	ABK46159	Kidney cancer rela
29	80.4	6.6	2400	22	ABK48579	CDNA encoding colo
30	80.4	6.6	2400	22	ABA66494	Human breast cell
31	80.4	6.6	2400	22	ABA33554	Human foetal liver
32	80.4	6.6	2400	22	AAK14917	Probe #12020 for g
33	80.4	6.6	2400	22	AAK10649	Human brain expres
34	80.4	6.6	2400	22	AA121411	Human bone marrow
35	80.4	6.6	2400	22	AA166636	Probe #11344 for g
36	80.4	6.6	2400	22	AA107105	Probe #15382 used t
37	80.4	6.6	2400	24	AB514609	Human genome-deriv
38	78.8	6.5	2981	22	AAK53131	Human polynucleoti
39	75.6	6.2	3135	22	AAK52147	Human colon cancer
40	74.6	6.2	1652	22	AAH34049	Human prostate exp
41	74.6	6.2	4017	23	ABV24327	Mouse Spt family t
42	74.2	6.1	1830	24	ABA86948	Human Spt family t
43	73.2	6.0	1197	23	ABA86949	Human Spt family t
44	67	5.5	1675	24	AA579321	DNA encoding novel
45	64.6	5.3	292	22	AA573932	DNA encoding zinc

ALIGNMENTS

RESULT 1
ID AAA15407 standard: DNA: 1213 BP.
XX
AC AAA15407:
XX
XX
DT 04-SEP-2000 (first entry)
XX
DE Fragment of DNA encoding a transcription factor designated htfi11A.
XX
KW Human; transcription factor; htfi11A: DNA-binding protein;
KW transcription; ribosomal RNA 5S gene; transcriptional control;
KW Cancer; ss.
XX
XX
OS Homo sapiens.
XX
PN WO200028024-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR02738.
XX
PR 10-NOV-1998; 98FR-0014146.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Rocher C;
XX
DR WPI; 2000-387419/33.
XX
PT New nucleic acid encoding human transcription factor htfi11A, useful for
treatment and diagnosis of cancer and inherited disease.

XX Claim 4: Page 42: 49pp: French.
 PS
 XX The present sequence represents a fragment of a human transcription
 CC factor (designated htfIIIA) gene. The polypeptide is probably a
 CC DNA-binding protein probably involved in initiating transcription of
 CC the gene for ribosomal RNA 5s and maintaining the stability of
 CC transcription of other control genes. The htfIIIA polynucleotides and
 CC polypeptides are used to make therapeutic or diagnostic compositions
 CC for diseases associated with disorders of transcriptional control,
 CC particularly cancer or other inherited diseases. The htfIIIA
 CC polynucleotide can also be used to detect anomalies in gene
 CC transcription, particularly for diagnosis of inherited disease, also
 CC for studying diseases involving htfIIIA.
 XX
 XX Sequence 1213 BP: 349 A; 309 C; 302 G; 253 T; 0 other:

Query Match 100.0%; Score 1213; DB 21; Length 1213;
 Best Local Similarity 100.0%; Pred. No. 2,2e-314;
 Matches 1213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGCGGCGCGCGCGAAGGTTGACAGGAGCGCGTGGCGCGCGCGGCGGTTCCCGC 60
 Db 1 GTGCGGCGCGCGCGGAAAGTTGACAGGAGCGCGTGGCGCGCGCGGCGGTTCCCGC 60
 QY 61 AGCTGTCTGCGCAGTGGCAGCGCGCTGGCCCTGGGCTTTGGAGCGCGCGCGCTTGA 120
 Db 61 AGCTGTCTGCGCAGTGGCAGCGCGCTGGCCCTGGGCTTTGGAGCGCGCGCGCTTGA 120
 QY 121 TCGCGCGCGCGTGTGCGCGAGTGTGTGTCTGTCAGCATGCGCGAGCGCTTCAATTC 180
 Db 121 TCGCGCGCGCGTGTGCGCGAGTGTGTGTCTGTCAGCATGCGCGAGCGCTTCAATTC 180
 QY 181 AGCGCGCGAGAGCTCAGTTCGACCGCGCGCGCGCGCGCGCTCCAGAGGTTCAATTC 240
 Db 181 AGCGCGCGAGAGCTCAGTTCGACCGCGCGCGCGCGCGCGCTCCAGAGGTTCAATTC 240
 QY 241 CTCTTCCCTTACTGTCAGCGCGCAATTACAGCAAAAGCTTGAAGCGCGCACCTGTG 300
 Db 241 CTCTTCCCTTACTGTCAGCGCGCAATTACAGCAAAAGCTTGAAGCGCGCACCTGTG 300
 QY 301 CAAAGCAACGCGGAGAGACCATTTGTTGTGACTATGAAGGTTGGCAAGCGCTTCAT 360
 Db 301 CAAAGCAACGCGGAGAGACCATTTGTTGTGACTATGAAGGTTGGCAAGCGCTTCAT 360
 QY 361 CAGGAGTACCATGTGACCGCGCAATTCGACTCAGACAGAGAAAGCCGTTTGTG 420
 Db 361 CAGGAGTACCATGTGACCGCGCAATTCGACTCAGACAGAGAAAGCCGTTTGTG 420
 QY 421 TCGACGCGCTGGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAACATTGGA 480
 Db 421 TCGACGCGCTGGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAACATTGGA 480
 QY 481 AGCGAAACATGAAATCAACAAATATATATGACGTTTGAAGACTTGAAGAGAC 540
 Db 481 AGCGAAACATGAAATCAACAAATATATATGACGTTTGAAGACTTGAAGAGAC 540
 QY 541 CTTTAAAGAAATGACGCGTGAATTCATGCGAGCATGACCAATGAAGCTCTATT 600
 Db 541 CTTTAAAGAAATGACGCGTGAATTCATGCGAGCATGACCAATGAAGCTCTATT 600
 QY 601 CAAAGTGTACCGAGGAAGATGTGGAAACATTTGATCACCAGCAAGCTGAAAGACA 660
 Db 601 CAAAGTGTACCGAGGAAGATGTGGAAACATTTGATCACCAGCAAGCTGAAAGACA 660
 QY 661 TCGCAAGGCGCGAGGCTATGTATGCAAAAAGATGTTCCTTGTGGCAAAAACATG 720
 Db 661 TCGCAAGGCGCGAGGCTATGTATGCAAAAAGATGTTCCTTGTGGCAAAAACATG 720
 QY 721 GACGGAACCTTGTGAACATGTGAGAGAAACCATTAAGAGGAATCTATGTGAAGTATG 780
 Db 721 GACGGAACCTTGTGAACATGTGAGAGAAACCATTAAGAGGAATCTATGTGAAGTATG 780

QY 781 CCGGAAAACATTTAAGCGAAAGATTACTTAAGCAACATGAAAACCTCATGCCCGACA 840
 Db 781 CCGGAAAACATTTAAGCGAAAGATTACTTAAGCAACATGAAAACCTCATGCCCGACA 840
 QY 841 AAGGATATATGCTGCTGTCCAGAGAGCGCTGTGAGAGACCTATCTACTGTGTTAA 900
 Db 841 AAGGATATATGCTGCTGTCCAGAGAGCGCTGTGAGAGACCTATCTACTGTGTTAA 900
 QY 901 TTTCAAAAGCCATATCTCTTCCTTCCATGAGAAACCGCCCTTTTGTGTGAACATGC 960
 Db 901 TTTCAAAAGCCATATCTCTTCCTTCCATGAGAAACCGCCCTTTTGTGTGAACATGC 960
 QY 961 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTCACTAGCATGCTGTGTACATGA 1020
 Db 961 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTCACTAGCATGCTGTGTACATGA 1020
 QY 1021 TCCCTACAAAGAAAGAAATGAAGCTCAAGCTCAAAAATCTCGTGAAGAAAGAGCTTGGC 1080
 Db 1021 TCCCTACAAAGAAAGAAATGAAGCTCAAGCTCAAAAATCTCGTGAAGAAAGAGCTTGGC 1080
 QY 1081 CTCTCATCTCAGTGTATATATCCCTCCCAAGGAAACAAAGGCAAGGCTTATCTTGTG 1140
 Db 1081 CTCTCATCTCAGTGTATATATCCCTCCCAAGGAAACAAAGGCAAGGCTTATCTTGTG 1140
 QY 1141 TCAAAAGGAGAGATCACCACTGTGTGAAAGCAAGATGCTGTGACAGTTGCACT 1200
 Db 1141 TCAAAAGGAGAGATCACCACTGTGTGAAAGCAAGATGCTGTGACAGTTGCACT 1200
 QY 1201 TACCTTGGCTAA 1213
 Db 1201 TACCTTGGCTAA 1213

RESULT 2
 AA15405
 ID AA15405 standard; DNA: 1273 BP.
 AC AA15405;
 XX
 DY 04-SEP-2000 (first entry)
 XX
 DE DNA encoding a human transcription factor designated htfIIIA.
 KW Human; transcription factor; htfIIIA; DNA-binding protein;
 KW transcription; ribosomal RNA 5S gene; transcriptional control;
 KW cancer; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 176..1273
 FT /*tag= a
 FT /product= "transcription factor"
 XX
 WO200028024-A1.
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99WO-FR02738.
 XX
 PR 10-NOV-1998; 98FR-0014146.
 XX
 PA (HMRI) HOECHST MARION ROUSSEL.
 XX
 PI Bordon-Pallier F. Rocher C;
 XX
 DR WPI: 2000-387419/33.
 DR P-PSDB: AAT93317.
 XX
 PT New nucleic acid encoding human transcription factor IIIA, useful for
 PT treatment and diagnosis of cancer and inherited disease
 XX
 PS Claim 3: Page 38-39; 49pp: French.

XX The present sequence encodes a human transcription factor designated hHflf1A. The polypeptide is probably a DNA-binding protein probably involved in initiating transcription of the gene for ribosomal RNA 5S and maintaining the stability of transcription of other control genes. CC The hHflf1A polynucleotides and polypeptides are used to make CC therapeutic or diagnostic compositions for diseases associated with CC disorders of transcriptional control, particularly cancer or other CC inherited diseases. The hHflf1A polynucleotide can also be used to CC detect anomalies in gene transcription, particularly for diagnosis of CC inherited disease, also for studying diseases involving hHflf1A. XX

SO Sequence 1273 BP; 356 A; 329 C; 328 G; 260 T; 0 other;

Query Match 99.9%; Score 1211.4; DB 21; Length 1273;

Best Local Similarity 99.9%; Prid No. 5.5e-314;

Matches 1212; conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 GTGCGGCGCGCGCGGAAAGTTACAGCAGAGCGGTGGCGCGCGCGGCTTCCCGC 60
DB 61 GTGCGGCGCGCGCGGAAAGTTACAGCAGAGCGGTGGCGCGCGCGGCTTCCCGC 120
OY 61 AGCTGTCTGGCAGCTGGCAGCGCGCGCGCGCTGGCGCTTGGAGGCGCGCGCGCTTGA 120
DB 121 AGCTGTCTGGCAGCTGGCAGCGCGCGCGCGCTGGCGCTTGGAGGCGCGCGCTTGA 180
OY 121 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 180
DB 181 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 240
OY 181 ACCCGCGAGACCTCAGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 240
DB 241 ACCCGCGAGACCTCAGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 300
OY 241 CTCCTTCCTTCTGACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 300
DB 301 CTCCTTCCTTCTGACTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 360
OY 301 CAAGCAGACGAGGAGAGACCATTTGTTGACTATGAGAGGCTGGCAGGCGCGCTTGC 360
DB 361 CAAGCAGACGAGGAGAGACCATTTGTTGACTATGAGAGGCTGGCAGGCGCGCTTGC 420
OY 361 CAGGAGCTACCTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 420
DB 421 CAGGAGCTACCTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGC 480
OY 421 TCGAGCACTGCGCTGTGATCAAAAAATCAACACAAATCAACCTTGAAGAACTTTGA 480
DB 481 TCGAGCACTGCGCTGTGATCAAAAAATCAACACAAATCAACCTTGAAGAACTTTGA 540
OY 481 AGCGAAACATGAAATCAACAAAAATATATATGAGCTTTGAAGAACTTTGAAGAA 540
DB 541 AGCGAAACATGAAATCAACAAAAATATATATGAGCTTTGAAGAACTTTGAAGAA 600
OY 541 CTTTAAAGAAATCATGAGCTGAAATTCATCATGCGACATACCATATGAACTTAT 600
DB 601 CTTTAAAGAAATCATGAGCTGAAATTCATCATGCGACATACCATATGAACTTAT 660
OY 601 CAAGGTATACCGAGGAAGATGTGGAAACATTGCTACCCGAGAACTGAAAGACA 660
DB 661 CAAGGTATACCGAGGAAGATGTGGAAACATTGCTACCCGAGAACTGAAAGACA 720
OY 661 TCGCAAGGCGCGCGAGGCTGTATGTCAAAAAAGATGCTTCTTGTGGCAAAAAAC 720
DB 721 TCGCAAGGCGCGCGAGGCTGTATGTCAAAAAAGATGCTTCTTGTGGCAAAAAAC 780
OY 721 GACGAACTCTGAAACATGTGAGAGAAACCATTAAGAGAAATCTATGAGAGTATG 780
DB 781 GACGAACTCTGAAACATGTGAGAGAAACCATTAAGAGAAATCTATGAGAGTATG 840
OY 781 CCGGAAACATTTAAACCAAGATTAACCTTAAGCAACATGAAACATCATGCCACAGA 840
DB 841 CCGGAAACATTTAAACCAAGATTAACCTTAAGCAACATGAAACATCATGCCACAGA 900

```

```

OY 841 AAGGATGTATGTGCTGTGCAAGAGAGCTGTGGAAGAACTATCTACTGTGTTAA 900
DB 901 AAGGATGTATGTGCTGTGCAAGAGAGCTGTGGAAGAACTATCTACTGTGTTAA 960
OY 901 TCTCCAAAGCATATCTCTCTCTTCCATGAGAAAGCGCGCTTTGTGTGACATGC 960
DB 961 TCTCCAAAGCATATCTCTCTCTTCCATGAGAAAGCGCGCTTTGTGTGACATGC 1020
OY 961 TGAGCTGTGCAAAACATTTGCAATGAACAAAGCTCCTAGAGCATCTGTGTACATGA 1020
DB 1021 TGAGCTGTGCAAAACATTTGCAATGAACAAAGCTCCTAGAGCATCTGTGTACATGA 1080
OY 1021 TCTGCAAGAGAAATGCAAGCTCAAAAGTCTGAGAAATCTGAGAAATCTGAGTGGC 1080
DB 1081 TCTGCAAGAGAAATGCAAGCTCAAAAGTCTGAGAAATCTGAGTGGC 1140
OY 1081 CTCTCATCTCACTGATATATCCCTCCCAAGAAACAGAGGCGCTTATCTTGTG 1140
DB 1141 CTCTCATCTCACTGATATATCCCTCCCAAGAAACAGAGGCGCTTATCTTGTG 1200
OY 1141 TCAAAAGGAGAGCTCAACCACTGTGTGAGAGCAACATGCTCTGACAGTTGCACTACT 1200
DB 1201 TCAAAAGGAGAGCTCAACCACTGTGTGAGAGCAACATGCTCTGACAGTTGCACTACT 1260
OY 1201 TACCTTGGCTAA 1213
DB 1261 TACCTTGGCTAA 1273

```

RESULT 3
AAH34586
ID AAH34586 standard; cDNA: 1349 BP.

AAH34586;
03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding cDNA SEQ ID NO:1668.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 13; ss.
OS Homo sapiens.
PN M0200122920-A2.
PN
PD 05-APR-2001.
PF 28-SEP-2000; 2000MO-US26524.
PR 29-SEP-1999; 990US-0157137.
PR 03-NOV-1999; 990US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCL INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA:
XX
XX WPI: 2001-235357/24.
XX P-PSDB: AAG75181.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 1; Page 3273-3274; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders

associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of p by expressing
 CC inactive proteins or to supplement the patients own production of p.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AH37196 to AH37204
 CC and AH37789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC XX

Sequence 1349 BP: 400 A; 337 C; 325 G; 285 T; 2 other:

Query Match 99.4%; Score 1206.2; DB 22; Length 1349;
 Best Local Similarity 99.8%; Pred. No. 1.3e-312;
 Matches 1208; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

3 GCCGGCGCCGCGCGAGGTTCAGCAGGAGCCGTGGCGCGCGCGCGCTTCCCGCAC 62
 DB GCGGGCGCGCGCGAGGTTTACAGGAGCCGTGGCGCGCGCGCGCTTCCCGCAC 81
 63 GTGTCTCGGCACGTGGAGCGCGCGCTGGAGGCGCGCGCGCGCTTGATC 122
 DB GTGTCTCGGCACGTGGAGCGCGCGCTGGAGGCGCGCGCGCGCTTGATC 141
 82 GTGTCTCGGCACGTGGAGCGCGCGCTGGAGGCGCGCGCGCGCTTGATC 141
 123 CGCGCGCGCGGTGGCGCGAGTGTGTCTGTCTTACCATGCGCGAGCGCTTATGCA 182
 DB CGCGCGCGCGGTGGCGCGAGTGTGTCTGTCTTACCATGCGCGAGCGCTTATGCA 201
 142 CGCGCGCGCGGTGGCGCGAGTGTGTCTGTCTTACCATGCGCGAGCGCTTATGCA 201
 183 CGCGCGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCTTCCAGAGAGTTCATCTGCT 242
 DB CGCGCGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCTTCCAGAGAGTTCATCTGCT 261
 202 CGCGCGAGAGCTCAGCTCGAGCCCGCGCGCGCGCGCTTCCAGAGAGTTCATCTGCT 261
 243 CCTTCCCTGACTGACAGCGCCCAATTACAGCAAAAGCTTGCAAGCGCGCACCTGTGCA 302
 DB CCTTCCCTGACTGACAGCGCCCAATTACAGCAAAAGCTTGCAAGCGCGCACCTGTGCA 321
 262 CCTTCCCTGACTGACAGCGCCCAATTACAGCAAAAGCTTGCAAGCGCGCACCTGTGCA 321
 303 AGCACACGGGAGAGACATTTGTTGTGACTGTGAAGGGTGTGCAAGGCGCTTATCA 362
 DB AGCACACGGGAGAGACATTTGTTGTGACTGTGAAGGGTGTGCAAGGCGCTTATCA 381
 322 AGCACACGGGAGAGACATTTGTTGTGACTGTGAAGGGTGTGCAAGGCGCTTATCA 381
 363 GGGACTACCATCTGAGCCGCCACATTTGACTACACAGAGAGAAAGCCGTTTGTGTC 422
 DB GGGACTACCATCTGAGCCGCCACATTTGACTACACAGAGAGAAAGCCGTTTGTGTC 441
 382 GGGACTACCATCTGAGCCGCCACATTTGACTACACAGAGAGAAAGCCGTTTGTGTC 441
 423 CAGCCACTGGCTGTGATCAAAAATTTCAACACAAATCAACTGAGAGAAACATTTGAAAC 482
 DB CAGCCACTGGCTGTGATCAAAAATTTCAACACAAATCAACTGAGAGAAACATTTGAAAC 501
 442 CAGCCACTGGCTGTGATCAAAAATTTCAACACAAATCAACTGAGAGAAACATTTGAAAC 501
 483 GCACAGATGAAAAATCAACAAAAAATATATATGCAATTTTGAAGACTGTAGAGAGCT 542
 DB GCACAGATGAAAAATCAACAAAAAATATATATGCAATTTTGAAGACTGTAGAGAGCT 561
 502 GCACAGATGAAAAATCAACAAAAAATATATATGCAATTTTGAAGACTGTAGAGAGCT 561
 543 TTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCCAGCATTAACCTTAATTTCA 602
 DB TTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCCAGCATTAACCTTAATTTCA 621
 562 TTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCCAGCATTAACCTTAATTTCA 621
 603 AGGTATCCAGAGAAAGATGTGGGAAACACTTTGATACCCAGCAAGCTGAAACGACATG 662
 DB AGGTATCCAGAGAAAGATGTGGGAAACACTTTGATACCCAGCAAGCTGAAACGACATG 681
 622 AGGTATCCAGAGAAAGATGTGGGAAACACTTTGATACCCAGCAAGCTGAAACGACATG 681
 663 CCAAGGCCACAGAGGCTATGTATGCAAAAAGGATGTTCTTTGTGCAAAAACATGGA 722
 DB CCAAGGCCACAGAGGCTATGTATGCAAAAAGGATGTTCTTTGTGCAAAAACATGGA 741
 682 CCAAGGCCACAGAGGCTATGTATGCAAAAAGGATGTTCTTTGTGCAAAAACATGGA 741
 723 CGGAACCTTGTGAACATGTGAGAGAAAGCCCTTAAGAGGAATATCTATGTGAAGTATGCC 782
 DB CGGAACCTTGTGAACATGTGAGAGAAAGCCCTTAAGAGGAATATCTATGTGAAGTATGCC 801
 742 CGGAACCTTGTGAACATGTGAGAGAAAGCCCTTAAGAGGAATATCTATGTGAAGTATGCC 801
 783 GGAACACATTTTAAACGCAAAAGATTACCTTAAGCAACATGAAGAACTCATGCCGAGAAA 842
 DB GGAACACATTTTAAACGCAAAAGATTACCTTAAGCAACATGAAGAACTCATGCCGAGAAA 861

DB 802 GCAAAACATTTTAAACGCAAAAGATTACCTTAAGCAACATGAAGAACTCATGCCGAGAAA 861
 843 GGGATGTATGTGCTGTGTCACAGAGAGGCTGTGGAAGAACCTATACCTATGTGTTAAATC 902
 DB GGGATGTATGTGCTGTGTCACAGAGAGGCTGTGGAAGAACCTATACCTATGTGTTAAATC 921
 862 GGGATGTATGTGCTGTGTCACAGAGAGGCTGTGGAAGAACCTATACCTATGTGTTAAATC 921
 903 TCCAAAGCATATCTCTCTCTCTCCATGATGAGAAAGCCGCTTTGTGTGAACATGCTG 962
 DB TCCAAAGCATATCTCTCTCTCTCCATGATGAGAAAGCCGCTTTGTGTGAACATGCTG 981
 922 TCCAAAGCATATCTCTCTCTCTCCATGATGAGAAAGCCGCTTTGTGTGAACATGCTG 981
 963 GCTGTGGCAAAACATTTGCAATGGAACAAAGTCTCAGTAGAGCAATGCTGTTGATGATC 1022
 DB GCTGTGGCAAAACATTTGCAATGGAACAAAGTCTCAGTAGAGCAATGCTGTTGATGATC 1041
 982 GCTGTGGCAAAACATTTGCAATGGAACAAAGTCTCAGTAGAGCAATGCTGTTGATGATC 1041
 1023 CTGACAGAGAAAGAAATGAGAGCTCAAAAGTCAAAAATCTGTGAAAAACGAGATTGGCCT 1082
 DB CTGACAGAGAGAAATGAGAGCTCAAAAGTCAAAAATCTGTGAAAAACGAGATTGGCCT 1101
 1042 CTGACAGAGAGAAATGAGAGCTCAAAAGTCAAAAATCTGTGAAAAACGAGATTGGCCT 1101
 1083 CTGATCTCAGTGTGATATATCTCTCCCAAAAGGAACAGGCGCAAGCTTATCTTGTGTC 1142
 DB CTGATCTCAGTGTGATATATCTCTCCCAAAAGGAACAGGCGCAAGCTTATCTTGTGTC 1161
 1102 CTGATCTCAGTGTGATATATCTCTCCCAAAAGGAACAGGCGCAAGCTTATCTTGTGTC 1161
 1143 AAACAGAGAGTCAACCCCAACTGTGTGGAAGACAAAGATGCTCTGACAGTTGCACTACTTA 1202
 DB AAACAGAGAGTCAACCCCAACTGTGTGGAAGACAAAGATGCTCTGACAGTTGCACTACTTA 1221
 1162 AAACAGAGAGTCAACCCCAACTGTGTGGAAGACAAAGATGCTCTGACAGTTGCACTACTTA 1221
 1203 CCTTGGCTTAA 1213
 DB CCTTGGCTTAA 1232

RESULT 4
 ABR83650
 ID ABR83650 standard; cDNA: 1381 BP.

AC ABR83650:
 DT 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #221.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

XX 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression

XX of genes associated with granulocyte activation, which serves as

XX PT diagnostic markers that is useful for monitoring disease states and

XX drug toxicity -
 XX Claim 1: SEQ ID No 221; 114pp; English.

CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in GC: (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, AIDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection and
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 1381 BP; 386 A; 350 C; 354 G; 291 T; 0 other;

XX Query Match 96.6%; Score 1171.6; DB 24; Length 1381;

XX Best Local Similarity 99.4%; Prod. No. 2.4e-303;
Matches 1207; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

XX 1 GTGCGGGGGCGCGGAAGGTTGACAGAGAGCGGTGGGGCGCGCGGCTTCCGGC 60
XX |||||||
XX 80 GTGCGGGCGCTCGCGGAAGGTTGACAGAGAGCGGTGGGGCGCGCGCGTTCGCGC 138
XX |||||||
XX 61 AGCTGTCGTGGGACGTGGAGCGGCGCTGGGCTTGGAGGGCGCGCGCCCTTGA 120
XX |||||||
XX 139 AGCTGTCTGGGACGTGGAGCGGCGCTGGGCTTGGAGGGCGCGCGCCCTTGA 198
XX |||||||
XX 121 TCGCGGGCGCGGTGGCGGAGTGGTGTCTTCCTTACCATGCGCGAGCTTCATTGC 180
XX |||||||
XX 199 TCGCGGGCGCGGTGGCGGAGTGGTGTCTTCCTTACCATGCGCGAGCTTCATTGC 258
XX |||||||
XX 181 AGCGGCGGAGACTTCAGTCCGAGCCCGCGCGCGCGCTTCCAGAGGTTTCATTCG 240
XX |||||||
XX 259 AGCGGCGGAGACTTCAGTCCGAGCCCGCGCGCGCGCTTCCAGAGGTTTCATTCG 318
XX |||||||
XX 241 CTCCTTCCTTGAATGAGCGCAATTACAGCAAAAGCTTGAAGCTTACCGGCACTGTG 300
XX |||||||
XX 319 CTCCTTCCTTGAATGAGCGCAATTACAGCAAAAGCTTGAAGCTTACCGGCACTGTG 378
XX |||||||
XX 301 CAAGCAGACGAGGAGAGACCATTTGTTGACTATGAAGGGTGGCAAGGCTTTCAT 360
XX |||||||
XX 379 CAAGCAGACGAGGAGAGACCATTTGTTGACTATGAAGGGTGGCAAGGCTTTCAT 438
XX |||||||
XX 361 CAGGAGCTACATCTGAGCGCGCACTTCTGACTACACAGAGAAAAGCCGTTGTTTG 420
XX |||||||
XX 439 CAGGAGCTACATCTGAGCGCGCACTTCTGACTACACAGAGAAAAGCCGTTGTTTG 498
XX |||||||
XX 421 TGCAGCAGCTGCTGATCAAAAATTCACACAAATCAACTGTAGAGAAATTTTGA 480
XX |||||||

DB 499 TGCAGCAATGCGCTGATCAAAAATTCACACAAATCAACTGTAGAGAAATTTTGA 558
DB 481 ACCGAACATGGAATAATCAACAAAAAATATATATGCGATTTTGAAGACTGTAAAGAC 540
DB 559 ACCGAACATGGAATAATCAACAAAAAATATATATGCGATTTTGAAGACTGTAAAGAC 618
DB 541 CTTAAGAAATCAACGAGCTGTAATAATTCATCATGTGCGACATTCACATGAACTTATT 600
DB 619 CTTAAGAAATCAACGAGCTGTAATAATTCATCATGTGCGACAAATTAACCTTATT 678
DB 601 CAAGTTACCGAGAGAGATGTGGAAACCTTTGCAATCCACGACGAAAGTGAAGACA 660
DB 679 CAAGTTACCGAGAGAGATGTGGAAACCTTTGCAATCCACGACGAAAGTGAAGACA 738
DB 661 TCGCAAGCGCGCAGAGGCTATGATATGTCACAAAAGATTTCTTGTGGCAAAACATG 720
DB 739 TCGCAAGCGCGCAGAGGCTATGATATGTCACAAAAGATTTCTTGTGGCAAAACATG 798
DB 721 GAGGAACCTTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 780
DB 799 GAGGAACCTTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 858
DB 781 CCGGAACATTTAAACGAAGATTAACCTTAACGACATGAAATCTATGCCCGACA 840
DB 859 CCGGAACATTTAAACGAAGATTAACCTTAACGACATGAAATCTATGCCCGACA 918
DB 841 AAGGATGATGTGCTGTGCAAGAGAGCTGTGGAAGAACTATCTACTGTCTTAA 900
DB 919 AAGGATGATGTGCTGTGCAAGAGAGCTGTGGAAGAACTATCTACTGTCTTAA 978
DB 901 TCTCCAAACGATTCCTCTCTCCATGAGAAAAGCGCGCTTGTGTGTAACATGC 960
DB 979 TCTCCAAACGATTCCTCTCTCCATGAGAAAAGCGCGCTTGTGTGTAACATGC 1038
DB 961 TGGCTGTGGCAAAATTTGCAATGAACAAAGTCTCCTAGCAGCATCTGTGTACATGA 1020
DB 1039 TGGCTGTGGCAAAATTTGCAATGAACAAAGTCTCCTAGCAGCATCTGTGTACATGA 1098
DB 1021 TCTGCAAGAGAAATGAACGTCGAAGTCAAAATTCGTAAGAAAC-GGAGTTTGG 1079
DB 1099 TCTGCAAGAGAAATGAACGTCGAAGTCAAAATTCGTAAGAAACGGAGTTTGG 1158
DB 1159 CCTCTCATCTCAGTGGATATAT-CCTCCAAAGAGAAACAGGCAAGCTTATCTTTGT 1217
DB 1140 GTCAAAACGAGAGTCAACCAACGTCGTGGAAGACAGATGCTCGACATGCGAGTAC 1199
DB 1218 GTCAAAACGAGAGTCAACCAACGTCGTGGAAGACAGATGCTCGACATGCGAGTAC 1277
DB 1200 TTACCTTGGCTAA 1213
DB 1278 TTACCTTGGCTAA 1291

RESULT 5
AAT14038
ID AAT14038 standard; cDNA: 1399 BP.
XX
AC AAT14038;
XX
DT 07-JUL-1996 (first entry)
XX
DE Transcription factor-11A gene.
XX
KW Human: transcription factor-11A; hTF11A; DNA binding protein;
KW ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE;
KW primer; PCR; polymerase chain reaction; foetal brain;
KW anchor primer; diagnostic probe; transcription control;
XX
XX antitumor; cancer; therapy; ss.
OS Homo sapiens.
XX

```

FH Key Location/Qualifiers
FT CDS 20..1291
FT /tag="a
FT /product="Transcription factor-111A
FT /note="ClalM 2"
FT misc-feature 317..1096
FT /tag="b
FT /product="zinc finger domains
FT polyA_signal 1363..1368
FT /tag="c
XX
XX EP704526-A1.
XX
XX 03-APR-1996.
XX
XX 05-SEP-1995; 95EP-0113908.
XX
XX 05-SEP-1994; 94JP-0211022.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
XX
XX Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;
XX MPI; 1996-173033/18.
XX
XX P-PSDB; AAR91305.
XX
XX Human Transcription Factor III A gene - useful in regulation of
XX transcription and for diagnosis and treatment of e.g. cancer related
XX diseases
XX
XX Example 1: Page 11-13; 17pp; English.
XX
XX The sequence encodes human transcription factor-111A (hTF111A), a
XX DNA binding protein with 9 zinc finger domains, which is necessary
XX for the initiation of 5S RNA gene transcription. Binding to an
XX internal control region of the 5S gene. The coding region
XX (claimed) is given in AAT14037. A fragment lacking a 5'-portion of
XX the gene has been isolated from a human foetal brain cDNA library
XX (OTK-1), and the 5'-portion of the gene has been isolated by
XX 5'-rapid amplification of cDNA ends using primers H11-R (AAT14039),
XX H11-E (AAT14040), H11-H (AAT14041) and AP-2 (AAT14042), complementary to
XX anchor primer AAT14043. Reverse transcription using H-11R is
XX followed by anchor primer ligation and PCR using AP-2 and H11-E, to
XX give a full-length cDNA, OTK7. The gene and its encoded protein
XX may be used in diagnosis, identification or therapy of hereditary
XX diseases such as cancer, or other diseases resulting from abnormal
XX transcriptional control, and to analyse the mechanisms involved in
XX their activity.
XX
XX Sequence 1399 BP; 405 A; 349 C; 354 G; 291 T; 0 other:
SQ
Query Match 96.5% Score 1170: DB 17: Length 1399:
Best Local Similarity 99.3% Pred. No. 6,5e-303:
Matches 1206: Conservative 0; Mismatches 5; Indels 3; Gaps 3:
QY 1 GTGCCGCCGCCGCCGGAAGTTTCAGACAGCGAGCGGTGGCCGGCCGCCCGTCCCGGC 60
DB 80 GTGCCGCCGCCGCCGGAAGTTTCAGACAGCGAGCGGTGGCCGGCCGCCCGTCCCGGC 138
QY 61 ACCTGTCTCGGACGTGGCAGCGCCGCTGGCCCTGGGCTTGGAGCGCGCGCCCTTGA 120
DB 139 ACCTGTCTCGGACGTGGCAGCGCCGCTGGCCCTGGGCTTGGAGCGCGCGCCCTTGA 198
QY 121 TCCGCGCGCGCGGTGTGCGCGAGTGTGTGTCTTTCAGCATGGCGAGCGCTTCATTGC 180
DB 199 TCCGCGCGCGCGGTGTGCGCGAGTGTGTGTCTTTCAGCATGGCGAGCGCTTCATTGC 258
QY 181 ACCGCGCGAGAGCTAGCTCCGACCGCGCGCGCGCGCGCTTCCAGAGAGTTTATCTG 240
DB 259 ACCGCGCGAGAGCTAGCTCCGACCGCGCGCGCGCGCGCTTCCAGAGAGTTTATCTG 318
QY 241 CTCCTTCCCTGACTGACGCGGCAATTACAGAAAGCTTGAGACTTGAGCGCCACTGTG 300

```

```

DB 319 CTCCTTCCCTGACTGACGCGGCAATTACAGAAAGCTTGAACACTTGAGCGCCACTGTG 378
QY 301 CAAGCACACGGGGGAGAGACGATTTGTTGTGACTATGAAGGGTGTGGCAAGCCCTTCAT 360
DB 379 CAGGACACGGGGGAGAGACGATTTGTTGTGACTATGAAGGGTGTGGCAAGCCCTTCAT 438
QY 361 CAGGACACTGACATCTGAGCGCGCACATTCGACTGCACAGAGGAAAGCCGTTGTTTG 420
DB 439 CAGGACACTGACATCTGAGCGCGCACATTCGACTGCACAGAGGAAAGCCGTTGTTTG 498
QY 421 TCGAGCCACTGGCTGTGATCAAAAATTCACAAATTCAAACTGAAAGAAACATTTTGA 480
DB 499 TCGAGCCAAATGGCTGTGATCAAAAATTCACAAATTCAAACTGAAAGAAACATTTTGA 558
QY 481 ACGAAACATGAAATTCACAAACAAATATATATATGCAAGTTTGAAGACTGTAAAGAC 540
DB 559 ACGAAACATGAAATTCACAAACAAATATATATATGCAAGTTTGAAGACTGTAAAGAC 618
QY 541 CTTTAAAGAAACATCAGCAGCTGAAATTCATCACTGCCAGCATACCAATGAACCTCTATT 600
DB 619 CTTTAAAGAAACATCAGCAGATGAAATTCATCACTGCCAGCATACCAATGAACCTCTATT 678
QY 601 CAAGTACCCAGAGAGGATGGGAAACCTTGCACTCCAGCAGCTGAAGACGACA 660
DB 679 CAAGTACCCAGAGAGGATGGGAAACCTTGCACTCCAGCAGCTGAAGACGACA 738
QY 661 TGCCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 720
DB 739 TGCCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAACATG 798
QY 721 GACGGAATCTTGAACATGTGAGAGAAACCATTAAGAGAAATCTTATGGAAGTATG 780
DB 799 GACGGAATCTTGAACATGTGAGAGAAACCATTAAGAGAAATCTTATGGAAGTATG 858
QY 781 CCGGAAACATTTAAAGCAAAAGATTACCTTAAAGCAACATGAATCATATGCCCGACA 840
DB 859 CCGGAAACATTTAAAGCAAAAGATTACCTTAAAGCAACATGAATCATATGCCCGACA 918
QY 841 AAGGATGTATGTCTGCTCCAGAGAGGCTGTGAAGAACTTACTACTGTGTTTAA 900
DB 919 AAGGATGTATGTCTGCTCCAGAGAGGCTGTGAAGAACTTACTACTGTGTTTAA 978
QY 901 TCTCCAAACCATATCTCTCTCTTCATGAGGAAAGCCGCTTTTGTGTGAAACATGC 960
DB 979 TCTCCAAACCATATCTCTCTCTTCATGAGGAAAGCCGCTTTTGTGTGAAACATGC 1038
QY 961 TGGCTGTGGCAAAACATTTGCAATGAACAAAGTCTCAGTAGGCATGCTGTGATCATGA 1020
DB 1039 TGGCTGTGGCAAAACATTTGCAATGAACAAAGTCTCAGTAGGCATGCTGTGATCATGA 1098
QY 1021 TCTGTCAAGAAAGAAATGAAGCTCAAAAGTCAAAAATCTCTGGAAGAAC -GGAAGTTGG 1079
DB 1099 TCTGTCAAGAAAGAAATGAAGCTCAAAAGTCAAAAATCTCTGGAAGAACGGAAGTTGG 1158
QY 1080 CCTCTCATCTCAGTGTATATCCCTCCAAAGGAAGGCAAGGCAAGGCTTATCTTGT 1139
DB 1159 CCTCTCATCTCAGTGTATATAT -CCTCCAAAGGAAGGCAAGGCTTATCTTGT 1217
QY 1140 GTCAAAACGAGAGTCAACCAACTGTGTGGAAGCAAGATGCTGTGACAGTTGCAGTAC 1199
DB 1218 GTCAAAACGAGAGTCAACCAACTGTGTGGAAGCAAGATGCTGTGACAGTTGCAGTAC 1277
QY 1200 TTACCTTTGGCTTAA 1213
DB 1278 TTACCTTTGGCTTAA 1291

```

RESULT 6
 AAT14037
 ID AAT14037 standard: cDNA: 1269 BP.
 XX
 AC AAT14037;
 XX

```

DT 07-JUL-1996 (first entry)
XX Transcription factor-IIIA gene.
DE
XX Human: transcription factor-IIIA; hTFIIIA; DNA binding protein;
XX ribosome; zinc finger; rapid amplification of cDNA ends: 5'-RACE;
KM primer: PCR: polymerase chain reaction; foetal brain;
KM anchor primer: diagnostic probe: transcription control;
KM antitumor; cancer; therapy; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX misc_feature 336..1115
XX FT /tag= a
XX FT /product= zinc finger domains
XX FT 1382..1387
XX FT polyA_signal /tag= b
XX
XX EP04526-A1.
XX
XX 03-APR-1996.
XX
XX 05-SEP-1995; 95SEP-0113908.
XX
XX 05-SEP-1994; 94JP-0211022.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
XX
XX Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;
XX WPI: 1996-173033/18.
XX P-PSDB: AAR91305.
XX
XX Human Transcription Factor III A gene - useful in regulation of
XX transcription and for diagnosis and treatment of e.g. cancer related
XX diseases
XX
XX Claim 2: Page 10-11; 17pp; English.
XX
XX The sequence encodes human transcription factor-IIIA (hTFIIIA),
XX a DNA binding protein with 9 zinc finger domains, which is necessary
XX for the initiation of 5S RNA gene transcription, binding to an
XX internal control region of the 5S gene. A fuller cDNA sequence
XX with flanking regions is given in AAT14038. A fragment lacking a
XX 5'-portion of the gene has been isolated from a human foetal brain
XX cDNA library (OTK7-1), and the 5'-portion of the gene has been
XX isolated by 5'-rapid amplification of cDNA ends using primers H11-R
XX (AAT14039), H11-E (AAT14040), H11-H (AAT14041) and Ap-2 (AAT14042).
XX Complementary to anchor primer AAT14043. Reverse transcription using
XX H-11R is followed by anchor primer ligation and PCR using Ap-2 and
XX H11-E, to give a full-length cDNA, OTK7. The gene and its encoded
XX protein may be used in diagnosis, identification or therapy of
XX hereditary diseases such as cancer, or other diseases resulting from
XX abnormal transcriptional control, and to analyse the mechanisms
XX involved in their activity.
XX
XX Sequence 1269 BP; 358 A; 324 C; 329 G; 258 T; 0 other:
XX
XX Query Match 96.2%; Score 1167; DB 17; Length 1269;
XX Best Local Similarity 99.3%; Pred. No. 4e-302;
XX Matches 1203; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```

```

Db 180 TCCGCGGCCGCTGCTCCCGAGTCGTCGTCTCTTACCATCGCCGAGCTTCAATTGC 239
Qy 181 AGCCGGGAGAGCTCAGCTCCGACCCCGCCCGCCGCTTCCGAGAGTTCATCTG 240
Db 240 AGCCGGGAGAGCTCAGCTCCGACCCCGCCCGCCGCTTCCGAGAGTTCATCTG 259
Qy 241 CTCCTTCCCTGACCTGACGCGCCCAATTACAGAAAGCTGGAACCTTGACCGGACCTG 300
Db 300 CTCCTTCCCTGACCTGACGCGCCCAATTACAGAAAGCTGGAACCTTGACCGGACCTG 359
Qy 301 CAAGCACAGGGGAGAGACCATTTTGTGTGATGATGAGAGGTGTGGCAAGCCTTCAT 360
Db 360 CAAGCACAGGGGAGAGACCATTTTGTGTGATGATGAGAGGTGTGGCAAGCCTTCAT 419
Qy 361 CAGGACTACCATCTGAGCGCCGCACATCTGACTACACAGAGAAACCGTTGTTTG 420
Db 420 CAGGACTACCATCTGAGCGCCGCACATCTGACTACACAGAGAAACCGTTGTTTG 479
Qy 421 TGCAGCACTGCTGTGATCAAAATTCACACAAATTCAAACTGAAAGAACTTTTGA 480
Db 480 TGCAGCACTGCTGTGATCAAAATTCACACAAATTCAAACTGAAAGAACTTTTGA 539
Qy 481 AGCCAAACATGAAATTCACAAACAAATATATATGAGTTTGAAGACTGTAAAGAC 540
Db 540 AGCCAAACATGAAATTCACAAACAAATATATATGAGTTTGAAGACTGTAAAGAC 599
Qy 541 CTTTAAAGAAACATCAGCAGCTGAAATTCATCAGTGCAGCATACCAATGAACTCTATT 600
Db 600 CTTTAAAGAAACATCAGCAGATGAAATTCATCAGTGCAGCATACCAATGAACTCTATT 659
Qy 601 CAAGTGTACCGAAGAGAGTGGGAAACATTTGCAATCCACGACCAAGCTGAAAGACA 660
Db 660 CAAGTGTACCGAAGAGAGTGGGAAACATTTGCAATCCACGACCAAGCTGAAAGACA 719
Qy 661 TGCAGAGCCCAAGAGGCTATGATGTCAAAAAGATGTTCTTGTGGCAAAAACATG 720
Db 720 TGCAGAGCCCAAGAGGCTATGATGTCAAAAAGATGTTCTTGTGGCAAAAACATG 779
Qy 721 GACGGAACCTTCAAAACATGTGAGAGAACCCATTAAGAGAAATCTATGAACTATG 780
Db 780 GACGGAACCTTCAAAACATGTGAGAGAACCCATTAAGAGAAATCTATGAACTATG 839
Qy 781 CCGGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAAACTCATGCCCA 840
Db 840 CCGGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAAACTCATGCCCA 899
Qy 841 AAGGATGTATGCTGCTGCTCCCAAGAGGCTGTGGAAGAAACCTATACATGCTTTTA 900
Db 900 AAGGATGTATGCTGCTGCTCCCAAGAGGCTGTGGAAGAAACCTATACATGCTTTTA 959
Qy 901 TCTCAAGACCATATCTCTCTCTTCATGAGAAAGCCGCTTTTGTGTGAAACATGC 960
Db 960 TCTCAAGACCATATCTCTCTCTTCATGAGAAAGCCGCTTTTGTGTGAAACATGC 1019
Qy 961 TGGCTGTGCAAAACATTTGCAATGAACAAGCTCAGTAGCATGCTTTTATACATGA 1020
Db 1020 TGGCTGTGCAAAACATTTGCAATGAACAAGCTCAGTAGCATGCTTTTATACATGA 1079
Qy 1021 TCTTGCAAGAAAGAAATGAGCTCAAGTCAAAAATCTCTGTGAAAAAC -GGAGTTTGG 1079
Db 1080 TCTTGCAAGAAAGAAATGAGCTCAAGTCAAAAATCTCTGTGAAAAACGGGAGTTTGG 1139
Qy 1080 CCTCTCATCTCACTGATATATATCTCTCCCAAAAGAAACAAAGGCAAGGCTTCTTTT 1139
Db 1140 CCTCTCATCTCACTGATATATAT -CCTCCCAAAAGAAACAAAGGCAAGGCTTCTTTT 1198
Qy 1140 GTCAAAAAGAGAGTCAACCAATCTGTGTGAAGCAAGATGCTCTGCAAGTTGACATC 1199
Db 1199 GTCAAAAAGAGAGTCAACCAATCTGTGTGAAGCAAGATGCTCTGCAAGTTGACATC 1258
Qy 1200 TTACCTTGGAC 1210
Db 1259 TTACCTTGGC 1269

```



```
PD 24-JAN-2001.
XX
XX
PF 27-JUN-2000; 2000CN-0116791.
XX
XX 27-JUN-2000; 2000CN-0116791.
XX
XX (NANF-) NANFANG RES CENT STATE HUMAN GENE GROUP.
XX
XX Yang Y, Xiao H, Kang B:
XX
XX WPI: 2001-282654/30.
XX
XX P-PSDB: AAB98214.
XX
XX Human translation initiation factor protein and its coding sequence -
XX
XX Claim 1: Page 17 (disclosure); 20pp; Chinese.
XX
XX The present invention describes a human eukaryotic initiation factor
XX (eIF), designated heif3, which is expressed in human dendritic cells.
XX heif3 has translation initiation factor activity. Also described in the
XX present invention are methods for the preparation and detection of the
XX heif3 protein and nucleotide sequences. The present sequence encodes
XX heif3, as given in the present invention.
XX
XX Sequence 1401 BP; 459 A; 281 C; 306 G; 355 T; 0 other:
XX
XX Query Match 20.3%; Score 246; DB 22; Length 1401;
XX Best Local Similarity 100.0%; Pred. No. 1,66-55;
XX Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 968 GGCACAAACATTGGCAATGAAACAAAGCTCAGCAGCATGCTGTGACATGATCTGAC 1027
XX 1401 GGCACAAACATTGGCAATGAAACAAAGCTCAGCAGCATGCTGTGACATGATCTGAC 1342
XX
XX 1028 AAGAAGAAATGAAGCTCAAGTCAAAAAATCTGTGAAAAACGAGATTGGCCTCTCAT 1087
XX 1341 AAGAAGAAATGAAGCTCAAGTCAAAAAATCTGTGAAAAACGAGATTGGCCTCTCAT 1282
XX
XX 1088 CTCAGTGGATATATCCCTCCCAAGAAAGCAAGGCAAGGCTTATCTTGTGTCAAAC 1147
XX 1281 CTCAGTGGATATATCCCTCCCAAGAAAGCAAGGCAAGGCTTATCTTGTGTCAAAC 1222
XX
XX 1148 GGAGAGTACCCCACTGTGTGGAAGACAGATGCTCTCGACAGTTGCACTTACCTT 1207
XX 1221 GGAGAGTACCCCACTGTGTGGAAGACAGATGCTCTCGACAGTTGCACTTACCTT 1162
XX
XX 1208 GCCTAA 1213
XX 1161 GCCTAA 1156
XX
XX Db
XX
XX RESULT 14
XX AAS73534
XX ID AAS73534 standard; CDNA: 1889 BP.
XX
XX AAS73534:
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #9338.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
```

```
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB: ABG09347.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1: SEQ ID NO 9338; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 1889 BP; 488 A; 453 C; 417 G; 531 T; 0 other:
XX
XX Query Match 19.5%; Score 236; DB 23; Length 1889;
XX Best Local Similarity 95.5%; Pred. No. 9e-53; 10; Indels 3; Gaps 3;
XX Matches 273; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
XX
XX 926 CATGAGAAACCCGCCCTTTGTGTGTGAACATGCTGGCTGTGGCAAAACATTTGCATG 985
XX 2 CATGAGAAACCCGCCCTTTGTGTGTGAACATG-TGCCTGTGGCAAAACA-TTGCAAT 59
XX
XX 986 AACCAAGTCTCAGTACGATGCGTGTGATGATGCTGACAGCAAGAAATGAAGCTC 1045
XX 60 AACCAAGTCTCAGTACGATGCTGTGTGATGATG-TCTGCGCAAGAAATGAAGCTC 118
XX
XX 1046 AAAGTCAAAAAAATCTGTGAAAAACGAGATTGGCCTCTCATCTCAGTGCATATATCCCT 1105
XX 119 AAAGTCAAAAAAATCTGTGAAAAACGAGATTGGCCTCTCATCTCAGTGCATATATCCCT 178
XX
XX 1106 CCCAAAGGAAACAAAGGCAAGGCTTATCTTGTGCAAAAGGAGAGTACCCCAAGCT 1165
XX 179 CCCAAAGGAAACAAAGGCAAGGCTTATCTTGTGCAAAAGGAGAGTACCCCAAGCT 238
XX
XX 1166 GTGAAGACCAAGATGCTCTCGACAGTTGCAAGTACTTACCTTGGCTAA 1213
XX 239 GTGAAGACCAAAATGCTCTCGAGTTCAGTACTTACCTTGGCTAA 286
XX
XX RESULT 15
XX ABK45059
XX ID ABK45059 standard; CDNA: 234 BP.
XX
XX ABK45059:
XX
XX 05-JUN-2002 (first entry)
```


DE cDNA encoding colon tumour protein, SEQ ID No 610.
XX
KM Human; colon tumour; vaccine; colon cancer; immunogenic;
KW Immunotherapy; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200212328-A2.
XX
PD 14-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-0524218.
XX
PR 03-AUG-2000; 2000US-223283P.
PR 28-MAR-2001; 2001US-279763P.
PR 29-JUN-2001; 2001US-302051P.
XX
PA (CORI-) CORIXA CORP.
XX
PI King GE, Meagher MJ, Xu J, Secrist H;
XX
DR WPI: 2002-241739/29.
XX
XX
PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
PS Claim 1; SEQ ID No 610; 147pp; English.
XX
XX The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK44450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
XX
SQ Sequence 234 BP; 83 A; 50 C; 53 G; 48 T; 0 other;

Query Match 19.3%; Score 234; DB 24; Length 234;
Best Local Similarity 100.0%; Pred. No. 1,1e-52;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 560 CTGAAATCCATCGATGCCAGCATACCAATGAACTCTATTCAAGTGTACCGAGAAAGA 619
DB 1 CTGAAATCCATCGATGCCAGCATACCAATGAACTCTATTCAAGTGTACCGAGAAAGA 60

QY 620 TGTGGAAACACTTTGCATCAACCCAGCAAGCTGAAACGACATGCCAAGGCCACGAGGGC 679
DB 61 TGTGGAAACACTTTGCATCAACCCAGCAAGCTGAAACGACATGCCAAGGCCACGAGGGC 120

QY 680 TATGTATGTCAAAAAGGATGTTCTTTGTGGCAAAAACATGACGGAATCTTGAAACAT 739
DB 121 TATGTATGTCAAAAAGGATGTTCTTTGTGGCAAAAACATGACGGAATCTTGAAACAT 180

QY 740 GTGAGAGAAACCATTAAGAGCAATATGTAAGTATGCCGAAACATTT 793
DB 181 GTGAGAGAAACCATTAAGAGCAATATGTAAGTATGCCGAAACATTT 234

Search completed: February 10, 2003, 12:56:26
Job time : 293.158 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 13:06:36 ; Search time 21 Seconds

(without alignments)
720.898 Million cell updates/sec

Title: US-09-831-426c-2
Perfect score: 2008

Sequence: 1 MDPFAVAVESVSLTIADAF.....ESPNCVEDKMLSTVAVLTLG 365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 98

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	1993	99.3	423	1	TF3A_HUMAN
2	1055.5	52.6	366	1	TF3A_XENLA
3	1047	52.1	335	1	TF3A_XENLA
4	1024	51.0	339	1	TF3A_XENLA
5	1022	50.9	339	1	TF3A_XENLA
6	816	40.6	322	1	TF3A_XENLA
7	556.5	27.7	365	1	P43_XENLA
8	553.5	27.6	365	1	P43_XENLA
9	417.5	20.8	803	1	ZN83_HUMAN
10	410.5	20.4	616	1	ZN83_HUMAN
11	409	20.4	336	1	ZG57_XENLA
12	409	20.4	799	1	ZN83_HUMAN
13	406	20.2	698	1	ZN83_HUMAN
14	404	20.1	821	1	ZN83_HUMAN
15	403	20.1	574	1	YF73_HUMAN
16	399	19.9	1167	1	Z208_HUMAN
17	398.5	19.8	626	1	Z143_HUMAN
18	398.5	19.8	751	1	Z237_HUMAN
19	395.5	19.7	453	1	Z237_HUMAN
20	395	19.7	803	1	Z237_HUMAN
21	394.5	19.6	803	1	Z237_HUMAN
22	393	19.6	489	1	Z237_HUMAN
23	393	19.6	689	1	Z237_HUMAN
24	392.5	19.5	337	1	Z237_HUMAN
25	391	19.5	738	1	Z237_HUMAN
26	386.5	19.2	759	1	Z237_HUMAN
27	386.5	19.2	898	1	Z237_HUMAN
28	386	19.2	913	1	Z237_HUMAN
29	385	19.2	626	1	Z237_HUMAN
30	383	19.1	538	1	Z237_HUMAN
31	382.5	19.0	280	1	Z237_HUMAN
32	382	19.0	645	1	Z237_HUMAN
33	381.5	19.0	535	1	Z237_HUMAN

34	381.5	19.0	734	1	ZN42_HUMAN	P28698	homo sapien
35	381	19.0	711	1	Z175_HUMAN	O94473	homo sapien
36	380.5	18.9	595	1	ZN85_HUMAN	O03923	homo sapien
37	380.5	18.9	682	1	ZN45_HUMAN	O02386	homo sapien
38	379	18.9	683	1	Y972_HUMAN	O92386	homo sapien
39	378	18.8	652	1	Y978_HUMAN	O94892	homo sapien
40	377.5	18.8	595	1	Z217_HUMAN	O96966	homo sapien
41	377	18.8	604	1	Z300_HUMAN	O96169	homo sapien
42	376.5	18.8	446	1	ZN70_HUMAN	O96066	homo sapien
43	376.5	18.8	472	1	Z1M3_HUMAN	O96066	homo sapien
44	376.5	18.8	570	1	ZN76_HUMAN	P36508	homo sapien
45	376.5	18.8	697	1	HKR1_HUMAN	P10072	homo sapien

ALIGNMENTS

RESULT 1

ID	TF3A_HUMAN	STANDARD:	PRT:	423 AA.
AC	Q92664	Q13097	Q12963	
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Transcription factor Iiia (Factor A) (Trilila).			
GN	GRF3A.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RC	MEDLINE=95309028; PubMed=7789179;			
RA	Arakawa H., Nagase H., Hayashi N., Ogawa M., Nagata M.,			
RA	Fujiwara T., Takahashi E., Shin S., Nakamura Y.;			
RT	"Molecular cloning, characterization, and chromosomal mapping of a			
RT	novel human gene (GRF3A) that is highly homologous to Xenopus			
RT	transcription factor Iiia.";			
RT	Cytogenet. Cell Genet. 70:235-238(1995).			
RL	[2]			
RP	SEQUENCE OF 61-423 FROM N.A.			
RP	TISSUE=Brain;			
RP	MEDLINE=95347600; PubMed=7622052;			
RA	Drew P.D., Nagle J.W., Canning R.D., Ozato K., Biddison W.E.;			
RA	Becker K.G.;			
RT	"Cloning and expression analysis of a human cDNA homologous to			
RT	Xenopus Trilila.";			
RT	Gene 159:215-218(1995).			
RL	[3]			
RP	CHARACTERIZATION.			
RP	MEDLINE=94342241; PubMed=8063702;			
RA	Moorefield B., Roeder R.G.;			
RT	"Purification and characterization of human transcription factor			
RT	Iiia.";			
RL	J. Biol. Chem. 269:20857-20865(1994).			
CC	-I- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF			
CC	APPROXIMATELY 50 BASES WITHIN THE 3S RNA GENES, IS REQUIRED FOR			
CC	CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO			
CC	BINDS THE TRANSCRIBED 5S RNA'S. MAY INITIATE TRANSCRIPTION OF THE			
CC	5S RIBOSOMAL RNA GENE AND MAINTAIN THE STABILITY OF TRANSCRIPTION			
CC	OF OTHER GENES.			
CC	-I- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-I- TISSUE SPECIFICITY: UROGENITAL.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			

DR EMBL: D32257: BAA06988.1: -
 DR EMBL: U20273: AAA75623.1: -
 DR EMBL: U14134: AAA21873.1: -
 DR HSSP: P03001: 1TF83.
 DR TRANSFAC: T04953: -
 DR Genew: HGNC:4662: GTF3A.
 DR MIM: 600860: -
 DR InterPro: IPR000822: Znf_C2H2.
 DR Pfam: PF00096: zfc-C2H2: 9.
 DR PRINTS: PRO0048: ZINC_FINGER.
 DR SMART: SM00355: ZNF_C2H2: 8.
 DR PROSITE: PS00028: ZINC_FINGER_C2H2_1: 7.
 DR PROSITE: PS0157: ZINC_FINGER_C2H2_2: 8.
 DR Transcription regulation: Zinc-finger: Metal-binding; DNA-binding;
 KW RNA-binding; Repeat; Nuclear protein.
 KM DOMAIN 98 359 ZINC_FINGERS.
 FT ZN_FING 98 122 C2H2-TYPE.
 FT ZN_FING 128 152 C2H2-TYPE.
 FT ZN_FING 158 183 C2H2-TYPE.
 FT ZN_FING 190 212 C2H2-TYPE (ATYPICAL).
 FT ZN_FING 220 244 C2H2-TYPE.
 FT ZN_FING 247 271 C2H2-TYPE.
 FT ZN_FING 275 297 C2H2-TYPE.
 FT ZN_FING 304 329 C2H2-TYPE.
 FT ZN_FING 335 359 C2H2-TYPE.
 FT ZN_FING 359 399 C2H2-TYPE.
 FT CONFLICT 214 214 N -> H (IN REF. 2).
 FT CONFLICT 221 221 K -> KL (IN REF. 2; AAA21873).
 FT CONFLICT 223 248 MISSING (IN REF. 2; AAA21873).
 FT CONFLICT 378 387 SLASHSLASH -> EFGSSSWIT (IN REF. 1).
 SO SEQUENCE 423 AA: 46847 MW: A627D064A43EB6F0 CRC64:

Query Match 99.3%; Score 1993; DB 1: Length 423;
 Best Local Similarity 99.2%; Pred. No. 8.5e-146;
 Matches 362; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPAAVAESVSLTADAFIAGESSATPRPALPRPFCSPDDCSANYSKAKIDAH 60
 DB 59 LDPAAVAESVSLTADAFIAGESSATPRPALPRPFCSPDDCSANYSKAKIDAH 118
 QY 61 LCKHTGERPVCYEGCGKAFIRDYHLSRHLITHTGERPVCYCAATGCDOKENTSNLKH 120
 DB 119 LCKHTGERPVCYEGCGKAFIRDYHLSRHLITHTGERPVCYCAATGCDOKENTSNLKH 178
 QY 121 FERKHEMOQOYICSEFDDCKTKPKKQOKIKQCQHTMEPLFKCQDECCGHNFSPEKIK 180
 DB 179 FERKHEMOQOYICSEFDDCKTKPKKQOKIKQCQHTMEPLFKCQDECCGHNFSPEKIK 238
 QY 181 RHANAHGVCYCGKGSFVAKTWTLLKRVETHEKEILCEVCOKRTFKKDYLNQMKHTHA 240
 DB 229 RHANAHGVCYCGKGSFVAKTWTLLKRVETHEKEILCEVCOKRTFKKDYLNQMKHTHA 298
 QY 241 PEROVCGCPREGGRTTYTWNLOSHILSFHESSRPVCEHAGCGKTFAMQSLTRAVV 300
 DB 299 PEROVCGCPREGGRTTYTWNLOSHILSFHESSRPVCEHAGCGKTFAMQSLTRAVV 358
 QY 301 HDPKMKMKLVKSKREKSLASHLSGYIPPRKOGQGLSLCQNGESPNCVEDMLSTVA 360
 DB 359 HDPKMKMKLVKSKREKSLASHLSGYIPPRKOGQGLSLCQNGESPNCVEDMLSTVA 418
 QY 361 VLTLG 365
 DB 419 VLTLG 423

RESULT 2
 ID TF3A_XENLA STANDARD: PRT: 366 AA.
 AC P03001: 091856:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor Iiia (Factor A) (TFIIIA) (S-TFIIIA/O-TFIIIA).
 OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
 CC Xenopodinae; Xenopus.
 ON NCBI_TaxID:8335;
 RX SEQUENCE OF 23-366 FROM N.A.
 RX MEDLINE=85074456; PubMed=6210149;
 RA Ginsberg A.M., King B.O., Roeder R.G.;
 RT "Xenopus 5S gene transcription factor, TFIIIA: characterization of a
 RT cDNA clone and measurement of RNA levels throughout development.";
 RL Cell 39:479-489(1984).
 RN [1]
 RP SEQUENCE OF 23-366 FROM N.A.
 RP MEDLINE=86176722; PubMed=3754326;
 RA Yun Tso J., Van den Berg J., Korn L.J.;
 RT "Structure of the gene for Xenopus transcription factor TFIIIA";
 RL Nucleic Acids Res. 14:2187-2201(1986).
 RN [3]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 112-137.
 RX MEDLINE=86312888; PubMed=3755818;
 RA Taylor W., Jackson I.J., Siegel N., Kumar A., Brown D.D.;
 RT "The developmental expression of the gene for TFIIIA in Xenopus
 RT laevis.";
 RL Nucleic Acids Res. 14:6185-6195(1986).
 RN [4]
 RP SEQUENCE OF 1-35 FROM N.A.
 RX MEDLINE=89306603; PubMed=2744458;
 RA Scolto K.W., Kaullen H., Roeder R.G.;
 RT "Positive and negative regulation of the gene for transcription
 RT factor Iiia in Xenopus laevis oocytes.";
 RL Genes Dev. 3:651-662(1989).
 RN [5]
 RP SEQUENCE OF 1-26 FROM N.A.
 RX MEDLINE=91071587; PubMed=2253880;
 RA Kim S.H., Darby M.K., Joho K.E., Brown D.D.;
 RT "The characterization of the TFIIIA synthesized in somatic cells of
 RT Xenopus laevis.";
 RL Genes Dev. 4:1602-1610(1990).
 RN [6]
 RP REPEATS ANALYSIS.
 RX MEDLINE=85284956; PubMed=4040853;
 RA Miller J., McClachlan A.D., Kling A.;
 RT "Repetitive zinc-binding domains in the protein transcription factor
 RT Iiia from Xenopus oocytes.";
 RL EMBO J. 4:1609-1614(1985).
 RN [7]
 RP REPEATS ANALYSIS.
 RX MEDLINE=85231235; PubMed=4007166;
 RA Brown R.S., Sander C., Argos P.;
 RT "The primary structure of transcription factor TFIIIA has 12
 RT consecutive repeats.";
 RL FEBS Lett. 186:271-274(1985).
 RN [8]
 RP REPEATS ANALYSIS.
 RX Boehm S., Drescher B.;
 RT "Multiple internal repeats within the structure of the 5S RNA/DNA
 RT binding transcription factor TF-Iiia from Xenopus laevis.";
 RL Studia Biophys. 107:237-247(1985).
 RN [9]
 RP STRUCTURE BY NMR OF 10-101.
 RX MEDLINE=97397344; PubMed=9253405;
 RA Foster M.P., Wulke D.S., Radhakrishnan I., Case D.A.,
 RA Gottesfeld J.M., Wright P.E.;
 RT "Domain packing and dynamics in the DNA complex of the N-terminal
 RT zinc fingers of TFIIIA.";
 RL Nat. Struct. Biol. 4:605-608(1997).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 32-210.
 RX MEDLINE=98169475; PubMed=9501194;
 RA Nolte R.T., Conlin R.M., Harrison S.C., Brown R.S.;
 RT "Differing roles for zinc fingers in DNA recognition: structure of a
 RT six-finger transcription factor Iiia complex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:2938-2943(1998).

CC	-1-	FUNGEND:	ACTS AS BOTH A POSITIVE TRANSCRIPTION FACTOR FOR 5S RNA GENES AND A SPECIFIC RNA BINDING PROTEIN THAT COMPLEXES WITH 5S RNA IN OOCYTES TO FORM THE 7S RIBONUCLEOPROTEIN STORAGE PARTICLE.
CC	CC	MAY PLAY AN ESSENTIAL ROLE IN THE DEVELOPMENTAL CHANGE IN 5S RNA GENE EXPRESSION. INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF APPROXIMATELY 50 BASES WITHIN THE 5S RNA POLYMERASE III. ALSO CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO BINDS THE TRANSCRIBED 5S RNAs.	
CC	CC	-1-	SUBCELLULAR LOCATION: Nuclear.
CC	CC	-1-	ALTERNATIVE PRODUCTS: 2 ISOFORMS; S-TFIIIA (IN SOMATIC CELLS) AND O-TFIIIA (IN OOCYTES); ARE PRODUCED BY ALTERNATIVE INITIATION.
CC	CC	-1-	TISSUE SPECIFICITY: SYNTHESIZED IN OOCYTES AND, IN MUCH LOWER LEVELS, IN SOMATIC CELLS.
CC	CC	-1-	DEVELOPMENTAL STAGE: THE LEVELS FOLLOW THE TRANSCRIPTIONAL ACTIVITY OF OOCYTE TYPE 5S RNA GENES DURING EMBRYOGENESIS. PRESENT IN VERY HIGH LEVELS IN MATURING OOCYTES WHEN OOCYTE TYPE 5S GENES ARE BEING EXPRESSED, AND IN MUCH LOWER LEVELS IN SOMATIC CELLS WHERE THE OOCYTE TYPE GENES ARE NOT EXPRESSED.
CC	CC	-1-	PTM: THE N-TERMINUS IS BLOCKED.
CC	CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.lsb-slb.ch/announce/ or send an email to license@lsb-slb.ch).
DR	EMBL:	K02938;	AAA9967.1; -.
DR	EMBL:	X03681;	CAB51745.1; -.
DR	EMBL:	X03735;	CAB51745.1; JOINED.
DR	EMBL:	X03736;	CAB51745.1; JOINED.
DR	EMBL:	X03737;	CAB51745.1; JOINED.
DR	EMBL:	X03738;	CAB51745.1; JOINED.
DR	EMBL:	X03739;	CAB51745.1; JOINED.
DR	EMBL:	X15785;	CAB33786.1; -.
DR	PIR:	A03529;	TXW13
DR	PIR:	A24961;	A24961.
DR	PIR:	S40785;	S40785.
DR	PDB:	1TF3;	17-SEP-97.
DR	PDB:	1TF6;	08-JUL-98.
DR	TRANSFAC:	T02295;	-.
DR	InterPro:	IPR000832;	Znf_C2H2.
DR	Pfam:	PF000956;	zfc_C2H2_8.
DR	PRINTS:	PR00048;	ZINC-FINGER.
DR	SMART:	SM00355;	ZnFC2H2_9.
DR	PROSITE:	PS00028;	ZINC_FINGER_C2H2_1; 8.
DR	PROSITE:	PS01517;	ZINC_FINGER_C2H2_2; 7.
RW	Transcription regulation;	Zinc-finger;	Metal-binding; DNA-binding;
RW	RNA-binding; Repeat;	Nuclear protein;	Alternative initiation;
KW	Polymorphism;	3D-structure.	
FT	CHAIN	1	366
FT			TRANSCRIPTION FACTOR IITA, SOMATIC
FT			ISOFORM.
FT	CHAIN	23	366
FT			TRANSCRIPTION FACTOR IITB, OOCYTE
FT			ISOFORM.
FT	INIT_MET	23	23
FT	DOMAIN	35	298
FT	ZN_FING	35	59
FT	ZN_FING	65	89
FT	ZN_FING	95	120
FT	ZN_FING	127	151
FT	ZN_FING	157	181
FT	ZN_FING	184	210
FT	ZN_FING	214	236
FT	ZN_FING	243	268
FT	ZN_FING	274	298
FT	VARIANT	96	96
FT	VARIANT	319	319
FT	VARIANT	335	335
FT	VARIANT	356	356
FT	VARIANT	365	365
FT	VARIANT	365	365
FT	CONFLICT	4	4
FT	SEQUENCE	366 AA:	41911 MW; 292337E710292FEA24 CRC64;

```

Query Match: 52.6%; Score 1055.5; DB 1; Length 366.
Best Local Similarity 55.1%; Pred. No. 5.3e-74;
Matches 196; Conservative 46; Mismatches 107; Indels 7; Gaps 4;

Oy 5 AAVASVSSLTADAFIAGESSAPLPPRALPRFFICSPFDCSANYSKAMLDALHLCN 64
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 2 AAKVASVSSEADAGSLVTGEGMGERA--LPVVYKRYICISFDACGAANKMKQLQAHLCN 59
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Oy 65 TGEPFCVDYEGGGKAFIRDYHLNRILTHNGERPFVCAAGCGOFNKTSNLKKHEPK 124
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 60 TGEFPFPCKEEGEKGFTSLHHLTRLSLTHTGKNKFTCDSDSCDLRTTYTAAKKKHFNF 119
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Oy 125 HENQOKYITCFEDCKKTKRKHOOLKHOCOTHNPLEFKCTQBCCGKHFAFSPEKLKRAK 184
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 120 HNIKICYVCHFEFGCKAKRKNNOALKVHQFSHTOODPYECHECCKDRFLSPRLKRHK 179
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Oy 185 AHGYVCQK--GCSPAKTWTFELKHVRTRTHEETLEVCCKTEKKRDYLOKHQMHADE 242
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 180 VHAYGPCKKDSCDSFGKTTWTLYLKVAECHOQLAVCYOVRKRRHMDLRDHQKTHEKE 239
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Oy 243 KDVRCRCRECCGRTTYTVENLOSHTLSPIESRPVVCCHACCGTFAMKOSLTRAHVND 302
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 240 RTVYLCPDCCDDSYTLPAFNLSHSIQSFIEDRPVCEHCACGCKCFAMKKSLSRHISVVD 299
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Oy 303 PDKRMKLKVKRSREKSLASHSGYIPPKRGOGSLCONGSPVCVEDKMILST 358
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 300 PEKR--KLKEKCRPKRSLASRLTYGIYPKSKE--KNASVSGTERKTDLSLVKNKDSGT 352
      | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 3
TF3A_RANP1 STANDARD: PRF. 335 AA.
ID TF3A_RANP1
AC P34695;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Transcription factor IIA (Factor A) (PFIIA).
OS Rana pipiens (Northern leopard frog).
OC Embryoto; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
OX NCBI_TaxID=9404;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE=93013035; PubMed=1398134;
RA Gaekins C.J., Smith J.F., Ogilvie M.K., Hanas J.S.;
RT "Comparison of the sequence and structure of transcription factor
   IIA from Bufo americanus and Rana pipiens.";
RU Gene 120.197-206(1992).
CL -I- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
   APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
   CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE II. ALSO
   BINDS THE TRANSCRIBED 5S RNA S.
CC -I- SUBCELLULAR LOCATION: Nuclear.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: MS8211; AAA49534.1; -
DR EMBL: X58369; AAA1260.1; -
DR PIR: JC1441; JC1441.
DR HSSP: P03001; ITP3.
DR InterPro: IPRO00822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 9.
DR PRINTS: PR00048; ZINCfinger.
DR SMART: SM00355; ZnF_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
```


RESULT 5	TF3A_BUFAM	STANDARD:	PRC:	339 AA.
AC	P34694;			
DT	01-FEB-1994 (rel. 28, Created)			
DT	01-FEB-1994 (rel. 28, Last sequence update)			
DT	01-FEB-1995 (rel. 31, Last annotation update)			
DE	Transcription factor IIA (Factor A) (TFIIA).			
OS	Bufo americanus (American toad).			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumelostomi;			
CC	Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae;			
CC	Bufo.			
CC	NCBI_TaxID=8389;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Ovary;			
RX	MEDLINE=93013035; PubMed=1398134;			
RA	Gedkins C.J., Smith J.F., Ogilvie M.K., Hanas J.S.;			
RT	"Comparison of the sequence and structure of transcription factor			
RT	IIA from Bufo americanus and Rana pipiens.";			
RL	Gene 120:197-206(1992).			
CC	- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF			
CC	APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR			
CC	CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO			
CC	BINDS THE TRANSCRIBED 5S RNAs.			
CC	- SUBCELLULAR LOCATION: Nuclear.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: X58366; CAA1259.1; -.			
DR	PIR: JC1442; JC1442.			
DR	HSSP: P03001; 1T96.			
DR	Interpro: IPR000822; Znf_C2H2.			
DR	Pfam: PF00096; zf_C2H2; 9.			
DR	PRINTS: PR00048; ZINCfinger.			
DR	SMART: SMO0355; Znf_C2H2; 9.			
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.			
DR	PROSITE: PS0157; ZINC_FINGER_C2H2_2; 8.			
KW	Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;			
KW	RNA-binding; Repeat; Nuclear protein.			
KW	FT DOMAIN 12 274			
FT	ZINC FINGERS.			
FT	FT ZN_FING 12 36			
FT	C2H2-TYPE.			
FT	FT ZN_FING 42 66			
FT	C2H2-TYPE.			
FT	FT ZN_FING 72 97			
FT	C2H2-TYPE.			
FT	FT ZN_FING 104 128			
FT	C2H2-TYPE.			
FT	FT ZN_FING 134 158			
FT	C2H2-TYPE.			
FT	FT ZN_FING 161 187			
FT	C2H2-TYPE.			
FT	FT ZN_FING 190 212			
FT	C2H2-TYPE.			
FT	FT ZN_FING 219 244			
FT	C2H2-TYPE.			
FT	FT ZN_FING 250 274			
FT	C2H2-TYPE.			
SO	SEQUENCE 339 AA: 39500 MW: 24CE32569648308 CRC64:			
Query Match	50.9%; Score 1022; DB 1: Length 339;			
Best Local Similarity	56.1%; Pred. No. 1.8e-71;			
Matches 169; Conservative 41; Mismatches 97; Indels 10; Gaps				
OY	34 PALPREFICSPDCSANKWLDLHLCKHTGERPVCYDEGGCAFIINDYLSRHILT 93			
DB	6 PUYAKRFICSPDCSNATYKNNRKLQANHLCKHTGERPCTYEGEGKEGYVLHNLNHNVIS 65			
OY	94 HTGKRPVCAATGCDCKDCKFTKSNLKHNERHENDONOKOYICSPDCKKTKRKHOOLK 153			
DB	66 HTGKRPCKCTEKNCLATFTTASNMNLHKKRAHSSPAOVYCYVDFDCCOQPRKNQOLK 125			
OY	154 CQHTNPELEKCTOEGCGKHFASPSLKRKAHKAHGYCQKG--CSFYAKTWTELLKNVE 211			

Dd		126	YIHNNQPFKCEGGDCKCYPASPLKRNKHETNAGVPCGRKDSCTPEFGKYNNMNAE	185
Oy		212	THKEILLCEVCRTFFRKDYLKQHKHTAPEDVCRCREGCGRITTYTNLDSLSTR	271
Dd		186	LH-SEVCSICNTFERRKSFLKEHRKIIRBEIYVRCRENCDBTPTTKNLASHLTFPH	244
Oy		272	EESRFCEFIAGCGKTFPAKKOSILTEHAHVDPDDKKMKMLVKRSREKRSLASHSGYIP-	330
Dd		245	ENLPFCFECEGCGKTFPAKKOSIDLREHNTHDPEKKKM---VKPRPVRLSLSLSCYKPK	301
Oy		331	--FKRKGGLSLCONGPNCVCDKTLSTVAATLT	364
Dd		302	KSKKKKPSTPAMESQQPPASKADPVLLENLTL	338
RESULT 6				
TF3A_ICRP	TF3A_ICRP	STANDARD:	PRT:	322 AA.
AC	P79797;			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
Dd	Transcription factor IIfA (Factor A) (TFIIIA).			
Dd	Ictalurus punctatus (Channel catfish).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;			
OC	Ictaluriidae; Ictalurus.			
OX	NCBI_TaxId=798;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	TTSUB=Ovary;			
RC	MEDLINE=98086209; PubMed=9426540;			
RA	Ogilvie M.R., Hanes J.S.;			
RT	"Molecular biology of vertebrate transcription factor IIfA: cloning			
RT	and characterization of TfIIIA from channel catfish oocytes.";			
RL	Gene 203:103-112(1997).			
CC	-I- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF			
CC	APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES. IS REQUIRED FOR			
CC	CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO			
CC	BINDS THE TRANSCRIBED 5S RNAs (BY SIMILARITY).			
CC	-I- SUBCELLULAR LOCATION: Nuclear.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outpostation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as our content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
Dd	EMBL: Z68499; CAA32808.1; .			
DR	HSSP: P03001; ITf6.			
DR	InterPro: IPR000822; ZnF_C2H2.			
DR	Pfam: PF00096; zf-C2H2; 9.			
DR	PRINTS: PRO0048; ZINCINGER.			
DR	SMART: SMO0355; ZnF_C2H2_9.			
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.			
DR	PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 8.			
KW	Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;			
KW	RNA-binding; Repeat; Nuclear protein.			
KW	DOMAIN			
FT	ZN_FING	12	273	ZINC FINGERS.
FT	ZN_FING	12	36	C2H2-TYPE.
FT	ZN_FING	42	64	C2H2-TYPE.
FT	ZN_FING	70	95	C2H2-TYPE.
FT	ZN_FING	102	126	C2H2-TYPE.
FT	ZN_FING	132	156	C2H2-TYPE.
FT	ZN_FING	159	184	C2H2-TYPE.
FT	ZN_FING	188	211	C2H2-TYPE.
FT	ZN_FING	218	243	C2H2-TYPE.
FT	ZN_FING	249	273	C2H2-TYPE.
SO	SEQUENCE	322 AA;	37366 MW;	EC30CBEB58FB8456 CRC64;

```

Query Match Similarity   40.6% Score 816; DB 1; Length 322;
Best Local Similarity    46.3%; Pred. No. 1,1e-55;
Matches 146; Conservative 55; Mismatches 92; Indels 22; Gaps 5;

QY      38 RPTGSPDPCSNANYSKAMKLNDLCKNHGERPVCDEGGCGAKFIRDYHLSRHILTNIGE 97
       : :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
Db      10 KNPVCSFLNKKASFSKMKLEAHNYCNGHLRPFACD--RDCKTFCSTCOQLTRHNLSHSKG 67
        KPVCATAGCGDDGFNFKNLSLKPKFKRKNEONOKOYCISFEEDCKTKTFKHOOGLIHOCOT 157
        98 KPVGSLEDGESSEFIISTADLKNHYERVNQHHEKNYVCDYECCAKEFFPKKQOLSHNCSEM 127
        Db      68 KPYOGULEDGSGEFTIATGLKNHYERVNQHHENKVYCDYECCAKEFFPKKQOLSHNCSEM 127
        QY      158 NEPLFKCTQEGSCSKHPASPSSKLRPAKNAHEGYC-QKGCSEFAKTVELLVRETHKEE 216
           1: :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        Db      128 NQLPREFQVEYGECCKKYTTSKKLQKHEKNVDYPCAEBDCDFQGCMRYEVAHKKAHRDA 187
           1: :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        QY      217 ILCEVCRKTEFKFRKDYLCK-MKTAAPREDVDCRPREGCGRTTYTVFNLOSLHSFEESR 275
           :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        Db      188 LQDSACKAVBNKAMELFKKIKLRLVNLGVARYPRKCJKEGOQKLYTTHTFNLHLLSFEHGRI 247
           :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        QY      276 PIVCEAAGCCCTRAMQSITRAVVYNDPDKKMLKLYKSSEKRSLSLHSLGIPIPKRK- 334
           1: :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        Db      248 SFICPHDCGCCRAARMEGSLKRNAVVDQRKKLQKTKRCRRKK-----LEPRTNV 298
           1: :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        QY      335 -----QQCGTLST 341
           1: :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        Db      299 SDSELPAQHLGSLT 313
           1: :::: |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|

RESULT 7
P43_XENLA STANDARD;     PRG:    365 AA.
ID XENLA
AC P25456;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE P43 S5 RNA binding protein (42S P43) (Thesaurin B).
OS Xenopus laevis [Afrikan clawed frog].
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90253278; PubMed=2331751;
RA John K.E., Darby W.K., Crawford E.T., Brown D.D.;
RA "A finger protein structurally similar to Tefitla that binds
RT exclusively to 5S rRNA in Xenopus."
RL Cell 61:293-300(1990).
CL -1 FUNCTION: P43 IS A 5S RNA BINDING PROTEIN WHICH IS A MAJOR
CC CONSTITUENT OF OOCYTES AND COMPRESSES PART OF A 42S
CC RIBONUCLEOPROTEIN STORAGE PARTICLE.
CC -1 SUBUNIT: THE 42S RNP PARTICLE COMPRESES FOUR SUBUNITS EACH OF
CC WHICH CONTAINS ONE MOLECULE OF 5S RNA, THREE MOLECULES OF TRNA,
CC TWO MOLECULES OF P50 (EP1 ALPHA) AND ONE MOLECULE OF THE 5S RNA
CC BINDING PROTEIN 43.
CC -----
CC This Swiss-Pro entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way com-
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
Cc or send an email to license@isb-slb.ch).
Cc -----
DR EMBL: MJ2473; AAA49714.1; -.
DR PIR: C34895; C34895.
DR HSRP: P03001; ITP3.
DR InterPro: IPRO00822; znf.C2H2.
DR Pfam: PF00966; zf-C2H2_9.
DR PRINTS: PR00046; ZINCfinger.
DR SMART: SM00353; Znfc_C2H2_9.
DR ProSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
```

DR PROSITE P50157; ZINC_FINGER_C2H2_2; 7.
KW zinc-finger: Metal-binding: RNA-binding: Repeat.
FT DOMAIN 15 275 ZINC FINGERS.

FT ZN_FING 15 39 C2H2-TYPE.
FT FT 45 69 C2H2-TYPE.
FT ZN_FING 75 100 C2H2-TYPE.
FT ZN_FING 106 130 C2H2-TYPE.
FT ZN_FING 136 160 C2H2-TYPE.
FT ZN_FING 163 187 C2H2-TYPE.
FT ZN_FING 191 213 C2H2-TYPE.
FT ZN_FING 220 245 C2H2-TYPE.
FT ZN_FING 251 275 C2H2-TYPE.

SQ SEQUENCE 365 AA; 41694 MW; 2EB5BE6D7192995 CRC64:
Query Match 27.7%; Score 556.5; DB 1; Length 365;
Best Local Similarity 38.1%; Pred. No. 9,5e-36;
Matches 106; Conservative 47; Mismatches 122; Indels 3; Gaps 3;

CC 42 CSPDDCSANYSKMKLIDALHLCKHTGERPPVCDYBECSCAFLROYHLSRNLHTTGKKPV 101
DB 17 CPAGACAAFRACKBDCKLDHMHGSEKKPKCGIKDCDVFFRRKROLILNKRIHALKLKLS 76
OY 102 CAATCGDOAKFKNTSNLKHPERKENNOOKOYICSFEDCKTKFKKHODLKHOCQHNPL 161
DB 77 CPPTGCAKMTFSTFKSLSRRLKYKH-GEAVPLCKCFVPCKRSFRKKRALRHLSVHSNEPL 135
OY 162 FKCPDECGKHFPSPSKLRHAKAHGVYCO-KGSPYAKWTPLTLKNHRETKELICE 220
DB 136 SVCDPQCCKSSKSSVAKLVAHQHRHGFRGCEGOVAPSWLAQLQHNAK-HPLELOCA 194
OY 221 VCRRTFRKRDKLVOMHKMTAPREDVCORPREGCRTYTTFVNLOSHTLFHESSRPFFE 280
DB 195 ACKRPFKKASLRHKAHKTAKRKPLDLPCRPCDCKDTFSSENLTNHVRKLHLCLOTHRCR 254
OY 281 HAGCGKTPAKMOSLTRHAVVHDPRDKKKMLKYKKSREK 318
DB 255 HSGCTRFAMRESLLRHVLVDPERKKLKFVKGPSK 292

RESULT 8
P43_XENBO ID STANDARD PRI: 365 AA.
AC P25066:
DB 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE P43 5S RNA binding protein (42S P43) (Thesaurin B).
OS Xenopus borealis (kenyan clawed frog).
OU Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Embryota; Batrachia; Anura; Mesobatrachia; Pipridae;
OX Xenopodinae; Xenopus.
RN NCHI_TaxID=8354;
RP [1]
RX MEDLINE=90235278; PubMed=2331751;
RA "John K.E., Darby M.K., Crawford E.T., Brown D.D.;
RT A finger protein structurally similar to TFIIF that binds
RT exclusively to 5S RNA in Xenopus.";
RL Cell 61:293-300(1990).
CC -I- FUNCTION: P43 IS A 5S RNA BINDING PROTEIN WHICH IS A MAJOR
CC CONSTITUENT OF COCYTES AND COMPRISES PART OF A 42S
CC RIBONUCLEOPROTEIN STORAGE PARTICLE.
CC -I- SUBUNIT: THE 42S RNP PARTICLE COMPRISES FOUR SUBUNITS EACH OF
CC WHICH CONTAINS ONE MOLECULE OF 5S RNA, THREE MOLECULES OF TRNA,
CC TWO MOLECULES OF P50 (EPI-ALPHA) AND ONE MOLECULE OF THE 5S RNA
CC BINDING PROTEIN 43.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and that statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

or send an email to license@isb-sib.ch.

CC EMBL: M32471; AAA9712.1; -.
DR PIR: A34895; A34895.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 9.
DR PRINTS: PR00048; ZINCINGER.
DR SMART: SM00355; Znf_C2H2; 9.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
DR Zinc_finger: Metal-binding; RNA-binding; Repeat.
KW DOMAIN 15 275
FT ZN_FING 15 39 C2H2-TYPE.
FT ZN_FING 45 69 C2H2-TYPE.
FT ZN_FING 75 100 C2H2-TYPE.
FT ZN_FING 106 130 C2H2-TYPE.
FT ZN_FING 136 160 C2H2-TYPE.
FT ZN_FING 163 187 C2H2-TYPE.
FT ZN_FING 191 213 C2H2-TYPE.
FT ZN_FING 220 245 C2H2-TYPE.
FT ZN_FING 251 275 C2H2-TYPE.
SQ SEQUENCE 365 AA; 41726 MW; D7E8C76E80E42EF2 CRC64; .

Query Match 27.6%; Score 553.5; DB 1; Length 365;
Best Local Similarity 37.4%; Pred No. 1,6e-35;
Matches 108; Conservative 48; Mismatches 130; Indels 3; Gaps 3;

OY 31 PRRALPRRITCSPPDCSANYSKAWKLDLHCKHTGERPVCYDGGCGKAFIRYHLSRH 90
DB 6 PTEPSKQSVRCRPAAGCKAVYRKREGLRDHMGHSQKLMKGGKDCGKMFARROLOKH 65
OY 91 ILVTGKEKPEVCATGCDGKFKNTSKNLKHFKEKHEKNOQKQYICSEDECKTFPKHOLK 150
DB 66 MKRRLTLKKHSCTPAGCKMFFSTKSLSRKLYKH-GDAVPLKCVSPCKSKSFRRKALR 124
OY 151 THQGHNEBPLFKCTGGCGKHFASPSKLRNAKAHEGYVCO-KCCSVAAKTWELLKHV 209
DB 125 ITHSEHNEPLSYGDVPGCGKSTSAKLAHHRHNGYRCSEYECQCTVSEPTWALQTLH 184
OY 210 RETKEELLEVCKTEKTRKDYLDQMKHTHAPERDVCRCPRGCGRTYTFVFNLSHLS 269
DB 185 KK-HPLEQCAACKCKPRKKSALRRKATHAKNPLOLPCPRDCKKITSTYFNLTHLHK 243
OY 270 FHESSRPVCEHAGCGKTFAMKOSLTRNAVVDHDPKKMKLVKKSREK 318
DB 244 VHLGLQTHRCRHSNCTRSFAMRSLVRLVVDHPRKKLKLKFGRRPSK 292

RESULT 9
XZDB_HUMAN STANDARD; PRT; 803 AA.
ID XZDB_HUMAN
AC P96169; O9UBB3;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger X-linked protein XZDB.
GN XZDB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wiley P.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RS SEQUENCE OF 307-709 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94093547; PubMed=8268913;
RA Greig G.M., Sharp C.B., Carrel L., Willard H.F.;
RT "Duplicated zinc finger protein genes on the proximal short arm of
RT the human X chromosome: isolation, characterization and
RT X-inactivation studies.";

RL Hum. Mol. Genet. 2:1611-1618(1993).
CC -I- SUBCELLULAR LOCATION: Nuclear (potential).
CC -I- TISSUE SPECIFICITY: UBICUITOUS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AL031115; CAB36858.1; -.
DR EMBL: Z99130; CAB16205.1; -.
DR EMBL: L14788; AAC37522.1; -.
DR HSSP: P03001; 1PF3.
DR Genew: HGNC:13199; XZDB.
DR MIM: 300236; .
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; Zf-C2H2; 10.
DR SMART: SM00355; Znf_C2H2; 10.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 10.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 9.
KW Zinc_finger: Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT DOMAIN 15 20 POLY-GLY.
FT DOMAIN 83 89 POLY-GLY.
FT DOMAIN 271 572 ZINC FINGERS.
FT ZN_FING 271 295 C2H2-TYPE.
FT ZN_FING 304 328 C2H2-TYPE.
FT ZN_FING 334 358 C2H2-TYPE.
FT ZN_FING 364 386 C2H2-TYPE.
FT ZN_FING 393 417 C2H2-TYPE.
FT ZN_FING 424 448 C2H2-TYPE.
FT ZN_FING 454 478 C2H2-TYPE.
FT ZN_FING 484 508 C2H2-TYPE.
FT ZN_FING 514 538 C2H2-TYPE.
FT ZN_FING 547 572 C2H2-TYPE.
SQ SEQUENCE 803 AA; 84792 MW; FDS5A6D326FF5A5D4 CRC64;

Query Match 20.8%; Score 417.5; DB 1; Length 803;
Best Local Similarity 29.1%; Pred No. 9.8e-25;
Matches 111; Conservative 47; Mismatches 131; Indels 93; Gaps 14;

OY 18 DAFLAGESSAPRP-----PRPALPRR-----FICSPDCSANYSKAWKLD 58
DB 230 DLLLAEPAPAPAPAPAEAEAGPAALGPRGLSGGVLYLCEADCCGTFPKKHOLK 289
OY 59 AHICKNT---GERPVCYDGGCGKAFIRYHLSRHLITLTGSEKPEVCATGCDGKFKNTS 115
DB 290 VHLTHSSDQGRPFKPLGCGNTTTSYKLRHLSHDKLRPGLCPAEGCGSFTTVY 349
OY 116 NLKKNF-----ERKHEKNOQKQYICSEDECKTFPKHQQ 148
DB 350 NLKAMKHEQENSFKCEVSESPQAKLSAQRSHFPERPRYGCASFSGCKTFITVSA 409
OY 149 LKIHOCQHTNE-PLFKCTGGCGKHFASPSKLRNAKAHEG---VQCG-KCCSVAAKTWT 203
DB 410 LFSHRAHFRQDELFSCTSPGCKSKQYDKACRLKLTILHRSHTGRRPLLCFDDCGMNFMS 465
OY 204 ELKAVRETHKEELLEVCKTEKTRKDYLDQMKHTHAPERDVCRCPRGCGRTYTFVNL 263
DB 470 KILRH-----KRR-----HDDRRF-MCPVEGCGKSFTRAEHL 501
OY 264 OSHLSPHESSRPVCEHAGCGKTFAMKOSLTRNAVVDHDPKKMKLVK---KKSRE 317
DB 502 KGHST-ILGTRKPFVCCVAGCAGCAFSARSLYIHSKHLQDVDTWKSRCPISSCKNKLPTS 560
OY 318 KRSLASHLSGTIPPKRKQGL 339
DB 561 KHSMTM-----VRRHKVGDDL 578

RESULT 10

ZN93_HUMAN STANDARD: PRT: 616 AA.
 ID ZN93_HUMAN PRT: 616 AA.
 AC P35789; O912N8; Created)
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc finger protein 93 (zinc finger protein HNF34) (Fragment).
 GN ZNF93.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lameerdin J.E., McCready P.M., Skowronski E., Sakaladas G.,
 RA Burkhardt-Schultz K., Gordon L., Dias J., Scott D., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Dangnan L.,
 RA Erlar A., Christensen M., Georgescu A., Avila J., Altix C.,
 RA Andreise T., Trankhelm M., Amico-Keller G., Coefield J., Duarte S.,
 RA Lucas S., Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
 RA Sanders C., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 RA Carraro A.V.;
 RA "Sequence analysis of a 5.7 kb region in 19p13.1 between OLF and
 RT D19S455.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-191 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Belletfroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
 RA Marital J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
 CC -SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
 CC -SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AC007204; AAD22981.1; -;
 DR EMBL: M61873; AA83548.1; -;
 DR PIR: H39384; H39384.
 DR HSSP: P08046; 1A1H.
 DR Genew: H0NC13169; ZNF93.
 DR MIM: 603975; -;
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam: PF00096; zf-C2H2; 16.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR Prodom: PD000003; ZnF_C2H2; 7.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; ZnF_C2H2; 16.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 16.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 17.
 KW Transcription regulation; zinc finger; DNA-binding; Metal-binding;
 KW Nuclear protein; Repeat.
 FT NON_TER 1 1
 FT DOMAIN <1 71 KRAB
 FT ZN_FING 141 163 C2H2-TYPE (DEGENERATE).
 FT ZN_FING 169 191 C2H2-TYPE.
 FT ZN_FING 197 219 C2H2-TYPE.
 FT ZN_FING 225 247 C2H2-TYPE.
 FT

FT ZN_FING 253 275 C2H2-TYPE.
 FT ZN_FING 281 303 C2H2-TYPE.
 FT ZN_FING 309 331 C2H2-TYPE.
 FT ZN_FING 337 359 C2H2-TYPE.
 FT ZN_FING 365 387 C2H2-TYPE.
 FT ZN_FING 393 415 C2H2-TYPE.
 FT ZN_FING 421 443 C2H2-TYPE.
 FT ZN_FING 449 471 C2H2-TYPE.
 FT ZN_FING 477 499 C2H2-TYPE.
 FT ZN_FING 505 527 C2H2-TYPE.
 FT ZN_FING 533 555 C2H2-TYPE.
 FT ZN_FING 561 583 C2H2-TYPE.
 FT ZN_FING 589 611 C2H2-TYPE.
 FT CONFLICT 71 71 S -> SGP (IN REF. 2).
 FT CONFLICT 85 85 Q -> H (IN REF. 2).
 FT CONFLICT 115 115 S -> R (IN REF. 2).
 FT CONFLICT 132 132 C -> S (IN REF. 2).
 FT CONFLICT 149 149 V -> D (IN REF. 2).
 SO SEQUENCE 616 AA; 70572 MW; DF309833AB61160A CRC64;
 Query Match 20.4%; Score 410.5; DB 1; Length 616;
 Best Local Similarity 30.5%; Pred. No. 2.6e-24;
 Matches 109; Conservative 48; Mismatches 137; Indels 63; Gaps 17;
 Oy 40 FICSPPCSNYSKAKKLDALHCKHTGGERFYCDYEGCGKAFIDYILHSRIILHTGKRP 99
 Db 197 YICE--ECGKAFYSSALNTHKRIHTGKRPYCD--KCDKAFIASSSTLSKRIHTGKRP 252
 Oy 100 FVCAATG-----CDQKNTSKNLKHKFERKHNQKQYI 133
 Db 253 YKCECGKAFNQSSTLTNKKIKHTGKRPYKCECGKAFNQSSTLTNKK--KKIHTGKRPYV 310
 Oy 134 CSPFDCKTKTKKHOQKIKHOCTNPELFKCTOEGCGKAFNPSKLRHKAHNG---YV 190
 Db 311 C--ECCGKAFYRIILTHKRIHTGKRPYCNK--CGKAFIASSSTLSNHEIHTGKRNK 366
 Oy 191 CQK-GCSPVAKTTELKIKVETNKEE--ILCEVGRKTPKRRDYLOKHKTAHERDVC 247
 Db 367 CEECGKAFI---SSVLTNRKRVHTGKRPYKCECGKAFYSSSTLSHKRSHTGKRPYK 422
 Oy 248 CPREGCGRTTYTNFNLQSHLSFHESSRFVCEHNGCGKTFAMQSLTRHRAVVDPPKK- 306
 Db 423 C--ECCGKAFYASSSTLSKRII-ITCKKRYKCE--CGKAFNQSSTLTNKKIKHTGKRPY 477
 Oy 307 KMLKIVKSKHEKNSLASHLSGYIPKRGCGSLICQN---GSPNCEVDKMLST 358
 Db 478 KCECGKAFNQSSTLTNKKIHTGKRPYKCECGKAFNQSSTLTNKKIKHT 528
 RESULT 11
 ZG57_XENLA STANDARD: PRT: 336 AA.
 ID ZG57_XENLA PRT: 336 AA.
 AC P18728;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Gastrula zinc finger protein XLOC57.1 (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
 OC Xenopodidae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90040658; PubMed=2509712;
 RA Naeftfeld W., El-Baradi T., Mentzel H., Pieler T., Koester M.,
 RA Poetting A., Knoechel W.;
 RT "Second-order repeats in Xenopus laevis finger proteins.";
 RL J. Mol. Biol. 208:639-659(1989).
 DR PIR: S06578; S06578.
 DR HSSP: P25450; 12NM.
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam: PF00096; zf-C2H2; 12.
 DR

DR SMART: SM00355; ZNF_C2H2. 12.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 11.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 12.
 DR ZINC_Finger: Metal-binding; DNA-binding; Repeat.
 FT NON_TER 1 1
 FT ZN_FING 6 28 C2H2-TYPE.
 FT ZN_FING 34 56 C2H2-TYPE.
 FT ZN_FING 62 84 C2H2-TYPE.
 FT ZN_FING 90 112 C2H2-TYPE.
 FT ZN_FING 118 140 C2H2-TYPE.
 FT ZN_FING 146 168 C2H2-TYPE.
 FT ZN_FING 174 196 C2H2-TYPE.
 FT ZN_FING 202 224 C2H2-TYPE.
 FT ZN_FING 230 252 C2H2-TYPE.
 FT ZN_FING 258 280 C2H2-TYPE.
 FT ZN_FING 286 308 C2H2-TYPE.
 FT NON_TER 336 336
 SO SEQUENCE 336 AA; 815B667602330B46 CRC64;
 Query Match 20.4%; Score 409; DB 1; Length 336;
 Best Local Similarity 34.0%; Pred. No. 1.8e-24;
 Matches 108; Conservative 38; Mismatches 122; Indels 50; Gaps 15;
 Oy 40 FICSPPCSNVSNAMKLDHLCKHTGGERPVCDEGGCAKATIRPHYHLSRHILHTGCKP 99
 Db 34 FICT--BCGGFSGSKGLTLOHMKTHGCKPFC--TECGKNFAQITTLIRHLHTGCKP 89
 Oy 100 FVCAATGCCDCKFKTSMLKKNFERKHENOQKOYICSPEDCKKTKKHOOLKHOCCHTNE 159
 Db 90 FSC--TECGKHFHAKHGLVSHM--KTHGCKEPFCT--ECGKNFAQKHLVSHMKTHGCE 143
 Oy 160 PLFCTGCGGKHPASHPSKLRHAKAHG--YVCGK--GCSVVAATWTELLKHAH-ETHK 214
 Db 144 KPFCTG--CGKNFAQKTNLCHLKITHTGCKPFTCEGCDKFAK--NLLIRHLKITHTGE 199
 Oy 215 EELICEVCKTFKRDYLOKHMKTAPEDVCRCPREGCCRTYTVFNLQSHLSFHEES 274
 Db 200 KPFCTGCGAFLTKGSLVGHMKITHGCKPFC--CGKNFTQKNSLCH-LLTMHTGE 255
 Oy 275 RPFVCEHAG-----CGKTFAMKQSLTHAVVHPDCKKM 308
 Db 256 KPFCTGCGGAFKAGMLVHTKITHGCKPFCSTCGCKNFAQKNSLHLKITHTREKFT 315
 Oy 309 KLVK-KSKREKSLASHL 325
 Db 316 YSECGKRYSOIVNLASHM 333
 RESULT 12
 ZKXA_HUMAN STANDARD; PRT; 799 AA.
 ID ZKXA_HUMAN STANDARD; PRT; 799 AA.
 AC P96166; Q9UUP7; 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE zinc finger X-linked protein ZKXA.
 GN ZKXA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bird C.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 303-759 FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=94093547; PubMed=8268913;
 RA Greig G.M., Sharp C.B., Carrel L., Willard H.F.;
 RT "Duplicated zinc finger protein genes on the proximal short arm of
 RT the human X chromosome: isolation, characterization and
 RT X-inactivation studies.";

RL Hum. Mol. Genet. 2:1611-1618(1993).
 CC -I- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -I- TISSUE SPECIFICITY: UBICUITOUS.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AL034396; CAB46717.1; .
 DR EMBL; L14787; AAC37521.1; .
 DR HSSP; P03001; 1TF3.
 DR Genew; HGNC:13198; ZKXA.
 DR MIM; 300235; .
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 10.
 DR SMART; SM00355; ZNF_C2H2. 10.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 9.
 DR ZINC_Finger: Metal-binding; DNA-binding; Repeat; Nuclear protein.
 FT DOMAIN 15 20 POLY-GLY
 FT ZN_FING 267 568 ZINC_FINGERS.
 FT ZN_FING 267 291 C2H2-TYPE.
 FT ZN_FING 300 324 C2H2-TYPE.
 FT ZN_FING 330 354 C2H2-TYPE.
 FT ZN_FING 360 382 C2H2-TYPE.
 FT ZN_FING 389 413 C2H2-TYPE.
 FT ZN_FING 420 444 C2H2-TYPE.
 FT ZN_FING 450 474 C2H2-TYPE.
 FT ZN_FING 480 504 C2H2-TYPE.
 FT ZN_FING 510 534 C2H2-TYPE.
 FT ZN_FING 543 568 C2H2-TYPE.
 FT CONFIDCT 753 759 AMSVHP -> SGVERTS (IN REF. 2).
 SO SEQUENCE 799 AA; 84770 MW; 112BF6C6DCA670 CRC64;
 Query Match 20.4%; Score 409; DB 1; Length 799;
 Best Local Similarity 29.2%; Pred. No. 4.4e-24;
 Matches 115; Conservative 48; Mismatches 137; Indels 94; Gaps 15;
 Oy 4 PAVVAESVSLTIADATIAAGESSAPTPPRPAL-----PRFICSPD----- 46
 Db 217 PGDCPELRSDDLAE--PAPAPAPQDEAGLAAAGPGLGSGPVLYLCEPAL 273
 Oy 47 CSANYSKAMKLDHLCKHT--GERPVCDEGGCAKATIRPHYHLSRHILHTGCKPFC 103
 Db 274 CGQFPKAKHQLKMHLLTHSSGQGRPFKCPGLGCGWTFSTYKLRHQLSHDKLRPFCCP 333
 Oy 104 ATGCDQKFNKSNLKHF-----ERKHENOQKOYICSF 136
 Db 334 AEGCKSFYTVYVNLKAMKHEQENSFKEVCESEPTQAKLAGHQRSHFEPERYOCAP 393
 Oy 137 EDCKTFKPKHOOLKIHOCCHTNE-PLFKCTOEGCGNHFASPSKLRHAKAHG--YVCG 192
 Db 394 SGCKKTFITVALFSNRAHFRDQELFSCSPFCSSKOYKACRLKILHLSHTGERFLCD 453
 Oy 193 -KGCSEFAKVTTELKIVRETHKEBELICEVCKTKFERKQDYLOKHMKTAPEDVCRCPRE 251
 Db 454 FDGCGWNETSMKSLRH-----KRX-----HDDRRR- MCPVE 485
 Oy 252 GCGRTYTVFNLQSHLSFHEESRPFYCEAGCGKTFAMKQSLTHAVVH--DPDKMM 309
 Db 486 GCGKSFYTAELHKLKSHSTI-HLGTKRPFCYVAGCCAFSARSSLYHSHKHLQDVTDKSR 544
 Oy 310 LKV-----KSKREKSLASHLSGITPPKKOGGL 339
 Db 545 CPISSCNKLFYSKSHMKTIM-----VRRHKVQDL 574
 RESULT 13
 ZK34_HUMAN

ID 2234 HUMAN STANDARD: PRT: 698 AA.
 AC Q14588; Q9NS45; Q9NT77;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 15-JUN-2002 (Rel. 41; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE zinc finger protein 234 (zinc finger protein H2F4) (Fragment).
 GN ZNF234 OR ZNF269.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 ON NCBI_TaxID=9606;
 RX MEDLINE=95169271; PubMed=7865130;
 RA Abrink M., Aveskog M., Hellman L.;
 RT "Isolation of cDNA clones for 42 different Kruppel-related zinc finger
 RT proteins expressed in the human monoblast cell line U-937.";
 RL DNA Cell Biol. 14:125-136(1995).
 RN [3]
 RP SEQUENCE OF 7-698 FROM N.A.
 RA Kodyrajani V., Ge Y., Severin J., Krummel G.K., Grable L.,
 RA Krikstad E., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
 RT "Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
 RT gene cluster.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF28417; AAF34785.1; -
 DR EMBL: X78927; CA55527.1; -
 DR EMBL: AC074331; AAF88104.1; -
 DR HSP: P08047; ISP2.
 DR Genew: HONG:13027; ZNF234.
 DR MIM: 604750; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf_C2H2; 19.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PR00048; ZINCINGER.
 DR PRODOM: PD000003; Znf_C2H2; 5.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; ZNF_C2H2; 19.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 18.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 19.
 KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 8 90 KRAB.
 FT DOMAIN 91 175 ZINC_FINGERS.
 FT ZN_FING 176 674
 FT ZN_FING 176 198 C2H2-TYPE.
 FT ZN_FING 204 226 C2H2-TYPE.
 FT ZN_FING 232 254 C2H2-TYPE.
 FT ZN_FING 260 282 C2H2-TYPE.
 FT ZN_FING 288 310 C2H2-TYPE.

FT ZN_FING 316 338 C2H2-TYPE.
 FT ZN_FING 344 366 C2H2-TYPE.
 FT ZN_FING 372 394 C2H2-TYPE.
 FT ZN_FING 400 422 C2H2-TYPE.
 FT ZN_FING 428 450 C2H2-TYPE.
 FT ZN_FING 456 478 C2H2-TYPE.
 FT ZN_FING 484 506 C2H2-TYPE.
 FT ZN_FING 512 534 C2H2-TYPE.
 FT ZN_FING 540 562 C2H2-TYPE.
 FT ZN_FING 568 590 C2H2-TYPE.
 FT ZN_FING 596 618 C2H2-TYPE.
 FT ZN_FING 624 646 C2H2-TYPE.
 FT ZN_FING 652 674 C2H2-TYPE.
 FT CONFLICT 494 494 N -> S (IN REF. 3).
 FT CONFLICT 600 600 A -> E (IN REF. 3).
 FT NON_TER 698
 SQ SEQUENCE 698 AA; 80271 MW; 936D390DD2E6746B CRC64;
 Query Match 20.2%; Score 406; DB 1; Length 698;
 Best Local Similarity 30.9%; Pred. No. 6.5e-24;
 Matches 105; Conservative 42; Mismatches 119; Indels 74; Gaps 14;
 Oy 46 DCSANYSKAWKLDANLCKHTGEPFVCDYEGGKAFIDYHLSHRILHTGKRPVCAAT 105
 Db 236 ECGKGFNRSTLTIVHCKLHSGEKPYNC--EEGGRAFIASHLOEGRHITGKRFKDDT- 292
 Oy 106 GCDQKENTKSNLKKHFERKHENDQOYTCSPDCKTKTKHNOOLKHOCOTNEPLFCT 165
 Db 293 -CGNFRFRRLANNHC--MVHTGKPYKC--EDCGKPTCSSNLRIHQRYHTGKPYKC- 346
 Oy 166 OEGCGKHAFSPSKLKHANAHEG---YVCO---KQ-----CSFY 198
 Db 347 -EEGCKEFLPDSFOAHRRIRHIGKPYVCYKCGKFIYSSSFOAHOGVHTGKPYKNEC 405
 Oy 199 AKITTELLK-----VETIKKEELICEVCKTKRKDYLYKQMKHNAEBRDYCKRPFGCG 254
 Db 406 GKSRMKIHQVHLVHTGKPYKVCGRKAFROSSLYLKHKAHSVQKP-KC--EECG 462
 Oy 255 RYTYTVENLOSHLSPHEESRPVCEHAG-----CGKTF 288
 Db 463 QGFQSSRLDIDHL-IHTGKPYKCECGKGFNRADLKICRHITGKPYNCECGKVF 521
 Oy 289 AMKQSLTRHNVHDPKXKMKLKYKSRKSLASHLSGY 328
 Db 522 SQASHLTHQVHSGEK--PFKCECGKSFSSRAHLQNH 558
 RESULT 14
 ID ZN41_HUMAN STANDARD: PRT: 821 AA.
 AC P51814; Q9UWC4; Q96LEB; Q9UWV5; Q9UWV6; Q9UWV7; Q9UWV8; Q9UWV9;
 AC Q9UWV0; Q9UWV1;
 DT 01-OCT-1996 (Rel. 34; Created)
 DT 15-JUN-2002 (Rel. 41; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE zinc finger protein 41.
 GN ZNF41.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 ON NCBI_TaxID=9606;
 RX MEDLINE=99380103; PubMed=10449920;
 RA Rosati M., Franze A., Matarazzo M.R., Grimaldi G.;
 RT "Coding region intron/exon organization, alternative splicing and
 RT X-chromosome inactivation of the KRAB/FPB-domain-containing human zinc
 RT finger gene ZNF41.";
 RL Cytogenet. Cell Genet. 85:291-296(1999).
 RN [2]
 RP SEQUENCE OF 280-821 FROM N.A.
 RA MEDLINE=91244317; PubMed=2037297;
 RX Franze A., Archidiacono N., Rocchi M., Marino M., Grimaldi G.;

```

FT  VARSPPLIC 25 66 MISSING (IN ISOFORM 3).
FT  VARSPPLIC 53 60 MISSING (IN ISOFORM 6 AND ISOFORM 8).
FT  VARSPPLIC 1 128 MISSING (IN ISOFORM 5).
FT  VARSPPLIC 141 176 MISSING (IN ISOFORM 4).
FT  VARSPPLIC 141 176 MISSING (IN ISOFORM 7 AND ISOFORM 8).
FO  SEQUENCE 111 111 Q -> R (IN REF. 1: CAB53039).
SO  SEQUENCE 821 AA: 93728 MW: F4AB7808CA98AF13 CRC64;

Query Match 20.1%; Score 404; DB 1; Length 821;
Best Local Similarity 32.6%; Pred. No. 1,le-23;
Matches 106; Conservative 46; Mismatches 121; Indels 52; Gaps 16.

OY 42 CSPPDSANYSKAMKIDAHLCNHTGSEPPVCDYEGCGKATFIDYILSRHILHTGKPRV 101
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 511 CS-DCGSESTKQSLVHORIHTEGKPRIC--TECGVFPHRNLTLTHOKHTGKPRYM 566
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
OY 102 CAATGCDOKFNTSNLTKRFRKRNHNOOKYTCSPDDCKKFKKHQDLKHOCOPHTPE 161
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 567 CAE--CGKATQDSNLIRH--OKHTGKPRKCN--CGGKAFIKSKIKTHOKSHIGENH 620
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
OY 162 FKCTOEGCGRHPSPSKLRKNAKANEH---YVCK--GCSFYAKTWELKNVR-ETHKEE 216
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 621 YEC--KKCGKAFIOKSTLSVHORIHTEGKPRVYCDGCGKAFIOK--SHFIAMHRITGKEP 676
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
OY 217 ILCEVCKTFPKRKYDKOMKTHADER--DVCRCRPGCGRTYTVFYLQSHLSLSPHEER 275
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 677 YECSDCKCTKTKSLRVHOKIHTEGKPRNIC---AECKAFETRSMLITH--OKHTPREK 731
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
OY 276 PFCV-----EHAG-----CGKTFAMKOSLGRHAVYHDPDKKKK 309
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 732 PYECGDGKFTFKSRSLNHOKSHTEGKPRYECSSKCGKAFIOKATLSMHOIHITGKPRYAC 791
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
OY 310 LKVKSPREKRS-LASHLSGYIPPKR 333
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 792 TECGKATPDRSLNLIKHKHNGSEKR 816
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

RESULT 15
YE73_HUMAN
ID YE73_HUMAN STANDARD; PRT; 574 AA.
AC O9P255;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIAA1473 (Fragment).
GN KIAA1473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId:9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nucleolus (Probable).
CC -1- SIMILARITY: BELONGS TO THE KREPEL FAMILY OF G2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sb.ch).
CC -----

```

DR EMBL: AB040906; BAA95997.1; -
DR HSSP: P07248; 1ARD.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 13.
DR Pfam: PF01352; KRAB; 1.
DR PRINTS: PRO0048; ZINC_FINGER.
DR PRODOM: PD000003; Znf_C2H2; 2.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; Znf_C2H2; 13.
DR PROSITE: PS00805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 13.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT NON_TER 1
FT DOMAIN 15 86 KRAB.
FT DOMAIN 184 582 ZINC_FINGERS.
FT ZN_FING 184 582
FT ZN_FING 184 206 C2H2-TYPE.
FT ZN_FING 212 234 C2H2-TYPE.
FT ZN_FING 240 262 C2H2-TYPE.
FT ZN_FING 268 290 C2H2-TYPE.
FT ZN_FING 296 318 C2H2-TYPE.
FT ZN_FING 324 346 C2H2-TYPE.
FT ZN_FING 352 374 C2H2-TYPE.
FT ZN_FING 380 402 C2H2-TYPE.
FT ZN_FING 408 420 C2H2-TYPE.
FT ZN_FING 436 458 C2H2-TYPE.
FT ZN_FING 464 486 C2H2-TYPE.
FT ZN_FING 492 514 C2H2-TYPE (DEGENERATE).
FT ZN_FING 520 542 C2H2-TYPE.
SQ SEQUENCE 574 AA; 65951 MW; 482E5F176FCASB0 CRC64;

Query Match 20.1%; Score 403; DB 1; Length 574;
Best Local Similarity 31.0%; Pred. No. 8.9e-24;
Matches 106; Conservative 47; Mismatches 121; Indels 68; Gaps 17;

QY 46 DCSANYSKAMLDLHLCCKHTGERPFVCDYEGGKAFINDYHLSRHLTHGCKPFVCAAT 105
DB 216 ECGKAYNETSLNTHKRIHTGKPKYC--EECGKAFNRLSHLTTHKRIHTGKPKYCEE- 272
QY 106 GCDOKFTKSNLKKHFERKHEND-----OKOYICSFED 138
DB 273 -CGKAFNQSANLTH--KRIHTGCKPKYCECGKAFSSOSTLTANKIITHGCKPKYC--EE 328
QY 139 CKKTFKKHQQLKIHOCQHTNEPLFKCTQEGCGKHFASPSKLRIHAKAHG---YVCGKGC 195
DB 329 CGKAFSSOSTLTTHKRIHTGCKPKYC--EECGKAFNRLSHLTTHKRIHTGCKPKYCEE-C 385
QY 196 SFVAKTTELKLVHRETHKEELI--CEVCGKTFKRKRDYIKOHMKTHAPRDVCGRPRGCG 254
DB 386 GKAFKSSSTLTTHKRIHAGCKPKYCEVCGKAFSRSHLTTHKRIHTGCKPKY--KC--EECG 442
QY 255 RTVYTVENLQSHLS--FHEESRPVCEHACCGKTFAMKOSLTRHAVVHDPDKKKMLK 311
DB 443 K---AFNLSSQLTTHKRIHTGCKPKYCEE--CGKAFNQSSTLSKHKVIHTGCK---PYK 493
QY 312 VKKREKRSLASHLGGYIPPKRKQSGSLSCNGESP--NCWE 352
DB 494 YECCGKAFNQSSTLTTH-----KMIHTGCKPKYCEE 524

Search completed: February 10, 2003, 17:45:49
Job time : 26 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:37:10 ; Search time 50.257 Seconds

(without alignments)
7401.927 Million cell updates/sec

Title: US-09-831-426C-4

Perfect score: 1213

Sequence: 1 gtgcgcgcgcgcgcgaag.....cagctactaccttgctaa 1213

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents -NA: *
1: /cgn2_6/ptodata1/jna/5A_COMB.seq: *
2: /cgn2_6/ptodata1/jna/5B_COMB.seq: *
3: /cgn2_6/ptodata1/jna/6A_COMB.seq: *
4: /cgn2_6/ptodata1/jna/6B_COMB.seq: *
5: /cgn2_6/ptodata1/jna/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata1/jna/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1170	96.5	1399	1 US-08-523-376-2	Sequence 2, Appli
2	1167	96.2	1269	1 US-08-523-376-1	Sequence 1, Appli
3	63.2	5.2	1471	4 US-09-492-985-11	Sequence 11, Appli
4	61.4	5.1	1430	4 US-09-492-985-1	Sequence 1, Appli
5	38.8	4.8	2992	4 US-09-362-123A-3	Sequence 3, Appli
6	34.8	4.5	3138	3 US-09-234-332-5	Sequence 5, Appli
7	53.2	4.4	2881	2 US-08-570-227A-1	Sequence 1, Appli
8	53.2	4.4	2881	4 US-09-077-991-1	Sequence 1, Appli
9	52.8	4.4	7218	4 US-08-232-463-14	Sequence 14, Appli
10	52.6	4.3	3600	4 US-09-657-042A-3	Sequence 3, Appli
11	51	4.2	4960	4 US-09-907-843-3	Sequence 3, Appli
12	50.4	4.2	1889	2 US-08-946-241B-1	Sequence 1, Appli
13	50.4	4.2	1889	2 US-08-946-241B-8	Sequence 8, Appli
14	50.4	4.2	1889	3 US-09-309-053-1	Sequence 1, Appli
15	50.4	4.2	1889	3 US-09-309-053-8	Sequence 8, Appli
16	50.4	4.2	2353	5 PCT-US92-06840-1	Sequence 1, Appli
17	47.8	3.9	2364	4 US-09-172-045-1	Sequence 1, Appli
18	43.6	3.6	2312	1 US-08-102-942A-1	Sequence 1, Appli
19	43.6	3.6	2312	4 US-09-037-179B-1	Sequence 1, Appli
20	43	3.5	563	4 US-09-385-982-433	Sequence 433, App
21	42	3.5	1680	1 US-08-234-783-3	Sequence 3, Appli
22	42	3.5	1680	1 US-08-456-907-3	Sequence 3, Appli
23	42	3.5	1680	5 PCT-US95-05523-3	Sequence 3, Appli
24	42	3.5	3090	1 US-08-102-942A-3	Sequence 3, Appli
25	41.4	3.4	2223	4 US-09-037-179B-3	Sequence 3, Appli
26	41.4	3.4	2223	1 US-08-317-522A-8	Sequence 8, Appli
27	41.4	3.4	2223	1 US-08-439-818A-8	Sequence 8, Appli

28	41.4	3.4	2223	2 US-08-751-965-8	Sequence 8, Appli
29	41.4	3.4	2223	2 US-08-738-975-8	Sequence 8, Appli
30	41.4	3.4	2223	2 US-08-728-626-8	Sequence 8, Appli
31	41.4	3.4	2223	3 US-08-808-599A-8	Sequence 8, Appli
32	41.4	3.4	2790	2 US-08-718-661-1	Sequence 1, Appli
33	40.8	3.4	4257	2 US-08-690-473-1	Sequence 1, Appli
34	40.8	3.4	4257	4 US-09-259-821A-1	Sequence 1, Appli
35	40.8	3.4	4257	4 US-08-843-653-1	Sequence 1, Appli
36	40.8	3.4	4405	1 US-07-885-972A-3	Sequence 3, Appli
37	40.8	3.4	4405	2 US-08-745-880-3	Sequence 3, Appli
38	40.8	3.4	4405	2 US-08-480-382-3	Sequence 3, Appli
39	40.8	3.4	12001	1 US-08-458-568A-11	Sequence 11, Appli
40	40.6	3.3	1892	2 US-08-933-750C-66	Sequence 66, Appli
41	40.6	3.3	1892	3 US-09-234-613-66	Sequence 66, Appli
42	40.2	3.3	530	3 US-08-758-663-4	Sequence 4, Appli
43	40.2	3.3	2888	4 US-08-765-907A-1	Sequence 1, Appli
44	40.2	3.3	4524	2 US-08-845-99A-7	Sequence 7, Appli
45	40.2	3.3	4524	3 US-09-206-537-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-523-376-2
; Sequence 2, Application US/08523376
; Patent No. 5808030
; GENERAL INFORMATION:
; APPLICANT: Tsumomu, FUJIWARA
; APPLICANT: Satoshu, TAKEDA
; APPLICANT: Yoshikazu, SHIMADA
; APPLICANT: Kouichi, OKAKI
; APPLICANT: Sadahito, SIN
; TITLE OF INVENTION: HFE11A GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,376
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; INFO: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: human fetal brain cDNA
; CLONE: OTK7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1288
; IDENTIFICATION METHOD: S
; US-08-523-376-2
Query Match 96.5%; Score 1170; DB 1; Length 1399;
Best Local Similarity 99.3%; Pred. No. 2.1e-303;

Matches 1206: Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```

QY 1 GTCCGCGCCGCCGGAAGTTTCAGCAGGAGCGCTGTGGCCGCGCGCGCTTCCGCGC 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 GTGCGCGCGCGGTGGGAGGTTTCAGCAGGAGCGCTGTGGCCGCGCGCGC-CGGTCTCCGCGC 138

QY 61 ACCTGTCTCGGACAGTGTGACAGCGCGCTGTGGCGCTTGTGAGGCGCGCGCGCTTGTGA 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139 ACCTGTCTCGGACAGTGTGACAGCGCGCTGTGGCGCTTGTGAGGCGCGCGCGCTTGTGA 198

QY 121 TCCGCGCGCGCTGTGTGCGCGAGTGTGTGTCTTGTGACCATCGCGAGCGCTTGTATTC 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 199 TCCGCGCGCGCTGTGTGCGCGAGTGTGTGTCTTGTGACCATCGCGAGCGCTTGTATTC 258

QY 181 ACCCGGCGAGAGCTCAGTCCGAGCCGCGCGCGCGCGCTTCCGAGGAGTTTCATCTG 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 259 ACCCGGCGAGAGCTCAGTCCGAGCCGCGCGCGCGCGCGCTTCCGAGGAGTTTCATCTG 318

QY 241 CTCCTTCCCTGACTGTGACGCGCAATTACAGCAAGCCTGTGAGCTTGTACGCGCAGCTGTG 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 CTCCTTCCCTGACTGTGACGCGCAATTACAGCAAGCCTGTGAGCTTGTACGCGCAGCTGTG 378

QY 301 CAGGACACGCGGGGAGAGACCATTTGTTGTGACTATGAAAGGTGTGGCAAGCCTTGTAT 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 379 CAGGACACGCGGGGAGAGACCATTTGTTGTGACTATGAAAGGTGTGGCAAGCCTTGTAT 438

QY 361 CAGGAGCTACATCTGAGCGCGCCCATTTGACTCACACAGAGAAAGCGCTTGTGTTG 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439 CAGGAGCTACATCTGAGCGCGCCCATTTGACTCACACAGAGAAAGCGCTTGTGTTG 498

QY 421 TCCAGCCAGCTGCTGTGATCAAAATTCACACAAATGAATGAATGAATGAATGAATGA 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 499 TCCAGCCAGTGGCTGTGATCAAAATTCACACAAATGAATGAATGAATGAATGAATGA 558

QY 481 ACCGAACATGAAATTCACAAACAAATATATATGACAGTTTGAAGCTGTAGAGAGAC 540
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 ACCGAACATGAAATTCACAAACAAATATATATGACAGTTTGAAGCTGTAGAGAGAC 618

QY 541 CTTTAAGAAATCAGCAGCTGAATATCATAGTSCAGCATACCATGAACTCTATTT 600
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 619 CTTTAAGAAATCAGCAGCTGAATATCATAGTSCAGCATACCATGAACTCTATTT 678

QY 601 CAAGGTACCCAGAGAGATGTGGAAACACTTTGCATCACCAGCAAGCTTAAACGACA 660
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 679 CAAGGTACCCAGAGAGATGTGGAAACACTTTGCATCACCAGCAAGCTTAAACGACA 738

QY 661 TCCCAAGGCCACGAGGCTATGTATGTCAAAAAGTGTCTTGTGTGCAAAAACATG 720
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 739 TCCCAAGGCCACGAGGCTATGTATGTCAAAAAGTGTCTTGTGTGCAAAAACATG 798

QY 721 GACGGAATTTGTGAAACACTGTGAGAGAAACCCATTAAGAGGAAATCTATGTGAAGTATG 780
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 799 GACGGAATTTGTGAAACACTGTGAGAGAAACCCATTAAGAGGAAATCTATGTGAAGTATG 858

QY 781 CCGGAAACATTTAAACGCAAGATATCTTAAGCAACATGAAACATCATGCCCGACA 840
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 859 CCGGAAACATTTAAACGCAAGATATCTTAAGCAACATGAAACATCATGCCCGACA 918

QY 841 AAGGATGTATGTGCTGTGTCCAGAGAGGCTGTGAGAGAACCTATACAGCTGTCTTTAA 900
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 919 AAGGATGTATGTGCTGTGTCCAGAGAGGCTGTGAGAGAACCTATACAGCTGTCTTTAA 978

QY 901 TCTCGAAAGCCATATCTCTCCCTTCATGAGAGAACCGCGCTTGTGTGCAACATGC 960
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 979 TCTCGAAAGCCATATCTCTCTTCATGAGAGAACCGCGCTTGTGTGCAACATGC 1038

QY 961 TGGCTGTGTGCAAAACATTTGTGATGAACAAGCTCTCAGTACGCTGTGTGATCAATGA 1020
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1039 TGGCTGTGTGCAAAACATTTGTGATGAACAAGCTCTCAGTACGCTGTGTGATCAATGA 1098

QY 1021 TCCGTGACAAAGAAATGAAGCTCAAGCTCAAAATATCTGTGAAAAAC-GGAGTTTGG 1079
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1099 TCCGTGACAAAGAAATGAAGCTCAAGCTCAAAATATCTGTGAAAAACGGAGTTTGG 1158

```

```

QY 1080 CCTTCATCTCAGTGTATATTCCTTCCCAAAAGAAAGGAGCGCATTCCTTTGT 1139
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1159 CCTTCATCTCAGTGTATATTCCTTCCCAAAAGAAAGGAGCGCATTCCTTTGT 1217

QY 1140 GTCAAAAGGAGAGTCAACCCCACTGTGTGAGAGCAAGAATGTCTTCGACAGTTGAGTAC 1199
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1218 GTCAAAAGGAGAGTCAACCCCACTGTGTGAGAGCAAGAATGTCTTCGACAGTTGAGTAC 1277

QY 1200 TTACCTTGTGCTAA 1213
   ||||| ||||| |||||
Db 1278 TTACCTTGTGCTAA 1291

RESULT 2
US-08-523-376-1
; Sequence 1, Application US/08523376
; Patent No. 5808030
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Satoshi, TAKEDA
; APPLICANT: Yoshikazu, SHIMADA
; APPLICANT: Kouichi, OZAKI
; APPLICANT: Sadahiro, SIN
; TITLE OF INVENTION: hTFIIIA GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,376
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-523-376-1

Query Match 96.2%; Score 1167; DB 1; Length 1269;
Best Local Similarity 99.3%; Pred. No. 1,3e-302;
Matches 1203: Conservative 0; Mismatches 5; Indels 3; Gaps 3;

```


Db	240	AGCCGGACGAGACCTAGCTCGACACCCGGCGCGCCCGCGCTTCCAGGAGGTTCACTCG	299
Oy	241	CTCTCTTCCCTGACTCTGACGGCCCAATTAACGCAAAACCTGGAAGCTTGAGCGCACTCTG	300
Db	300	CTCTCTTCCCTGACTCTGACGGCCCAATTAACGCAAAACCTGGAAGCTTGAGCGCACTCTG	359
Oy	301	CAGCAGCAACGGGGAGAGACCATTTGTGTGACTATGAAGGCTGGGCAAGCCCTCAT	360
Db	360	CAGCAGCAACGGGGAGAGACCATTTGTGTGACTATGAAGGCTGGGCAAGCCCTCAT	419
Oy	361	CAGGACATCACTGAGAGCCGCACATTTGACTACACAGAGAAAGCCGTTGTTTG	420
Db	420	CAGGACATCACTGAGAGCCGCACATTTGACTACACAGAGAAAGCCGTTGTTTG	479
Oy	421	TGCAGCCACTGGCTGTGATCAAAAATTCAACACAAATCAAACTTGAGAAACATTTTGA	480
Db	480	TGCAGCCACTGGCTGTGATCAAAAATTCAACACAAATCAAACTTGAGAAACATTTTGA	539
Oy	481	ACGCAAAACATGAAATTAACAAAAACAAATATATATGACGTTTAAAGCTGTAAGAAC	540
Db	540	ACGCAAAACATGAAATTAACAAAAACAAATATATATGACGTTTAAAGCTGTAAGAAC	599
Oy	541	CTTTTAAAGAAACATGACGCTGAAATTCATATAGTGCAGATACCAATGAACCTCTATT	600
Db	600	CTTTTAAAGAAACATGACGCTGAAATTCATATAGTGCAGATACCAATGAACCTCTATT	659
Oy	601	CAAGTGTACCGAGGAAGATGTGGGAAACATTTTGCATCACACCAGCAAGCTGAACGACA	660
Db	660	CAAGTGTACCGAGGAAGATGTGGGAAACATTTTGCATCACACCAGCAAGCTGAACGACA	719
Oy	661	TGCGAAGGGCCGACGGGCTATGTATGTCAAAAAGATGTCCTTTGTGGCAAAAAACATG	720
Db	720	TGCGAAGGGCCGACGGGCTATGTATGTCAAAAAGATGTCCTTTGTGGCAAAAAACATG	779
Oy	721	GACGGATCTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG	780
Db	780	GACGGATCTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG	839
Oy	781	CCGGAACCATTTTAAAGCAAAATTAACCTTAAGCAACATGAAACTACGCCACAGA	840
Db	840	CCGGAACCATTTTAAAGCAAAATTAACCTTAAGCAACATGAAACTACGCCACAGA	899
Oy	841	AAGGGATGATGTGGCTGTGCCAGAGAAAGCTGTGGAAGACATATACATCTGTGTTAA	900
Db	900	AAGGGATGATGTGGCTGTGCCAGAGAAAGCTGTGGAAGACATATACATCTGTGTTAA	959
Oy	901	TCTCCAAAGCCATATCCTCTCTTCCATGAGAAAGCCGCTTTGTGTGAACATG	960
Db	960	TCTCCAAAGCCATATCCTCTCTTCCATGAGAAAGCCGCTTTGTGTGAACATG	1019
Oy	961	TGCGTGTGGCAAAACATTTGCAATGAAGAAAGGTCTCACTAGGCATCTGTGTACATGA	1020
Db	1020	TGCGTGTGGCAAAACATTTGCAATGAAGAAAGGTCTCACTAGGCATCTGTGTACATGA	1079
Oy	1021	TCTGTGACAAAGAAATGAAGCTCAAAAGTCAAAAAATCTGTGTAAAAAC-GGAGTTTGG	1079
Db	1080	TCTGTGACAAAGAAATGAAGCTCAAAAGTCAAAAAATCTGTGTAAAAAGCGAGTTTGG	1139
Oy	1080	CGTCTCATCTCAGTGGATATATCCCTCCCAAAAGGAACAAGGCAAGCTTATCTTTGT	1139
Db	1140	CGTCTCATCTCAGTGGATATAT-CCTCCCAAAAGGAACAAGGCAAGCTTATCTTTGT	1198
Oy	1140	GTCAAAAGGAGAGTCAACCCACATCTGTGTGAAGACAAAGATCTCTGCACAGTTGCAATAC	1199
Db	1199	GTCAAAAGGAGAGTCAACCCACATCTGTGTGAAGACAAAGATCTCTGCACAGTTGCAATAC	1258
Oy	1200	TTACCTTGGC 1210	
Db	1259	TTTACCTTGGC 1269	

```

? Sequence 11 Application US/09492985
? Patent No 6376240
?
? GENERAL INFORMATION:
?
? APPLICANT: Song, An M.
? APPLICANT: Chen, Ya-Fen
? APPLICANT: Kremsky, Alan M.
? TITLE OF INVENTION: RFLAP-1: A Transcription Factor That
? FILE REFERENCE: SUN-113P
? CURRENT APPLICATION NUMBER: US/09/492,985
? CURRENT FILING DATE: 2000-01-27
? EARLIER APPLICATION NUMBER: 60/111,576
? EARLIER FILING DATE: 1999-01-27
? NUMBER OF SEQ ID NOS: 11
?
? SOFTWARE: FastSeq for windows Version 4.0
?
? SEQ ID NO 11
? LENGTH: 1471
?
? TYPE: DNA
? ORGANISM: mouse
?
? OS-09-492-985-11

```

Query Match	5.28;	Score 63.2;	DB 4;	Length 1471;
Best Local Similarity	48.98;	Pred. No. 2e-07;		
Matches 170; Conservative	0;	Mismatches 178;	Indels 0;	Gaps

[illegible]

RESULT 3
US-09-492-985-11

```

RESULT 4-92-985-1
US-09-492-985-1
: Sequence 1, Application US/09422985
: Patent No. 6376240
: GENERAL INFORMATION:
: APPLICANT: Song, An M.
: APPLICANT: Chen, Ya-Fen
: APPLICANT: Krensky, Alan M.
: TITLE OF INVENTION: RFLAN-1: A Transcription Factor That
: TITLE OF INVENTION: Activates RANRES Gene Expression
: FILE REFERENCE: SUN-113P
: CURRENT APPLICATION NUMBER: US/09/492,985
: CURRENT FILING DATE: 2000-01-27
: EARLIER APPLICATION NUMBER: 60/117,576
: EARLIER FILING DATE: 1999-01-27
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO. 1
: LENGTH: 1430
: TYPE: DNA
: ORGANISM: homo sapien
: FEATURE:
: OTHER INFORMATION: synthetic oligonucleotide probe

```



```

; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,227A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D
; REGISTRATION NUMBER: 30,440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-359-3260
; TELEFAX: 612-359-3263
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2881 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
;
; US-08-570-227A-1

```

```

Query Match          4.4%: Score 53.2; DB 2; Length 2881;
Best Local Similarity 54.6%: Pred. No. 0.00013;
Matches 106; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

QY 227 AGAGGTTTCATCTGCTCCCTGCTGACTGACGCCCAATTACAGCAAGCCGTGAAGCTT 286
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 1185 AGAGTCACATCTGTAGCCACCCAGAGTGCGCAAGACATCTTAAAGTTCCCATCTG 1244
      || || || || || || || || || || || || || || || || || || || ||
QY 287 GAGCGCCACTGTGACAGACAGCGGGAGAGACCATTTTGTGTGACTATGAAGGCTGT 346
      || || || || || || || || || || || || || || || || || || || ||
DB 1245 AAGGCCACACAGAGACCGACACAGAGAAAGCCCTTTCAGCTGTGGAAGGTTGT 1304
      || || || || || || || || || || || || || || || || || || || ||
QY 347 GCGAAGCCCTTCATCAGGAGACTACATCTGAGCCGCCACATTTCTGACTCACAGAGAA 406
      || || || || || || || || || || || || || || || || || || || ||
DB 1305 GAAGAGAGGTTTGCCCGTTTGATGATGAACTGTCCAGACACAGGGAACCCACAGGGT 1364
      || || || || || || || || || || || || || || || || || || || ||
QY 407 AAGCCGTTTGTGTG 420
      || || || || || ||
DB 1365 AAGAAATTTGCGTG 1378

```

```

RESULT 8
US-09-077-991-1
; Sequence 1, Application US/09077991
; Patent No. 6207375
; GENERAL INFORMATION:
; APPLICANT: Subramaniam, M.
; APPLICANT: Spelsberg, T.C.
; TITLE OF INVENTION: TGF-beta inducible early factor-1
; FILE REFERENCE: 150.157US2
; CURRENT APPLICATION NUMBER: US/09/077,991
; CURRENT FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: PCT/US96/19555
; EARLIER FILING DATE: 1996-12-11
; EARLIER APPLICATION NUMBER: US 08/570,227
; EARLIER FILING DATE: 1995-12-11
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0

```

```

; SEQ ID NO 1
; LENGTH: 2881
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; US-09-077-991-1

```

```

Query Match          4.4%: Score 53.2; DB 4; Length 2881;
Best Local Similarity 54.6%: Pred. No. 0.00013;
Matches 106; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

QY 227 AGAGGTTTCATCTGCTCCCTGCTGACTGACGCCCAATTACAGCAAGCCGTGAAGCTT 286
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 1185 AGAGTCACATCTGTAGCCACCCAGAGTGCGCAAGACATCTTAAAGTTCCCATCTG 1244
      || || || || || || || || || || || || || || || || || || || ||
QY 287 GAGCGCCACTGTGACAGACAGCGGGAGAGACCATTTTGTGTGACTATGAAGGCTGT 346
      || || || || || || || || || || || || || || || || || || || ||
DB 1245 AAGGCCACACAGAGACCGACACAGAGAAAGCCCTTTCAGCTGTGGAAGGTTGT 1304
      || || || || || || || || || || || || || || || || || || || ||
QY 347 GCGAAGCCCTTCATCAGGAGACTACATCTGAGCCGCCACATTTCTGACTCACAGAGAA 406
      || || || || || || || || || || || || || || || || || || || ||
DB 1305 GAAGAGAGGTTTGCCCGTTTGATGATGAACTGTCCAGACACAGGGAACCCACAGGGT 1364
      || || || || || || || || || || || || || || || || || || || ||
QY 407 AAGCCGTTTGTGTG 420
      || || || || || ||
DB 1365 AAGAAATTTGCGTG 1378

```

```

RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMUG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)835-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

```

CLONE: PTZapt-F18
US-08-232-463-14

Query Match 4.4%: Score 52.8; DB 1; Length 7218;
Best Local Similarity 4.5%: Pred. No. 0.00025;
Matches 18; Conservative 218; Mismatches 160; Indels 0; Gaps 0;

451 CACAAATCAATCTGAAGAAATTTGAAGCAACATGAAATCAACAAACAAATCA 510
1466 CAAGTAGTTAAAGAGATGAAGAATTGTCACRRRRRRRRRRRRRRRRRRRRRR 1407
511 TATATGAGTTTGAAGCTGTAAAGACCTTTAAGAAACATCAGCAGCTGAATCCA 570
1406 RRR 1347
571 TCAGTGCCGATACCATGAACCTTATTCAGTGTACCCAGGAAGATGTGGGAACA 630
1346 RRR 1287
631 CTTGCAATCACCAGCAAGCTGAAGACATGCCAGCCGACGAGGCTATGTATCA 690
1286 RRR 1227
691 AAAGAGATCTCTTTGGGCAAAACATGAGAGACCTCTGAACATGTGAGAAAC 750
1226 RRR 1167
751 CCATAAGAGCAATATCTATGTGAAGTATGCCGAAACATTTAAAGCAAGATTACCT 810
1166 RRR 1107
811 TAAGCAACATGAAGAACTATGCCCAAGCAAGAGA 846
1106 RRR 1071

RESULT 10
US-09-657-042A-3
Sequence 3, Application US/09657042A
Patent No. 6329203
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-1 EXPRESSION
FILE REFERENCE: RTS-0148
CURRENT APPLICATION NUMBER: US/09/657, 042A
CURRENT FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 88
SEQ ID NO 3
LENGTH: 3600
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (79)...(3399)
US-09-657-042A-3

Query Match 4.3%: Score 52.6; DB 4; Length 3600;
Best Local Similarity 51.2%: Pred. No. 0.0002;
Matches 149; Conservative 0; Mismatches 139; Indels 3; Gaps 1;

281 AAGCTTGACGGCCACTCTGCAGACACAGGGGAGAGACCAATTTGTGACTATGAA 340
937 ATGCTGTGCTTACATCGCAGACACATGGGAGAACCCACACAAAGTCACGTTTGA 996
341 GGGTGGCAAGGCTTATCATGAGGACTACATGCGCGGACAGATTTGACTGACACA 400
997 GGGTCCGGAAGTCACTACAGCCCTGCAAAACCTGGAAGCAGCACTGGGGTCACACAG 1056
401 GGAGAAAACCCGTTGTTTGACAGCCAGCTGGCTGTGATCAAAAATTCACAAATCA 460
1057 GGTGAGAAAGCCATACATGTGTGACAGAGGGGCTGACGTAAAGCCTTCAGCAATGCCAGT 1116

461 AACTTGACGAACATTTGAAAGCGCAACATGAAATCAACAAACAAATATATGCACT 520
1117 GACCGAGCCACACACAGAAATCGAGCCATTCCAAAT--GAGAAAGCGGTATGATGAAG 1173
521 TTGGAAGACTGTGAAGAGACCTTTAAGAAATCAGCAGCTGAAATTCAT 571
1174 CTCCTGGCTGCACCAACGCTATATGAGATTCATGCTGCTGGGAACAT 1224

RESULT 11
US-09-907-843-3
Sequence 3, Application US/09907843
Patent No. 6440739
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-2 EXPRESSION
FILE REFERENCE: RTS-0279
CURRENT APPLICATION NUMBER: US/09/907, 843
CURRENT FILING DATE: 2001-07-17
NUMBER OF SEQ ID NOS: 87
SEQ ID NO 3
LENGTH: 4960
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (572)...(4348)
US-09-907-843-3

Query Match 4.2%: Score 51; DB 4; Length 4960;
Best Local Similarity 52.5%: Pred. No. 0.00063;
Matches 136; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

281 AAGCTTGACGGCCACTCTGCAGACACAGGGGAGAGACCAATTTGTGACTATGAA 340
1052 ATGCTGTGCTTACATCGCAGACACAGGGGAGAACCCACAAAGTCACGTTTGA 1111
341 GGGTGGCAAGGCTTATCATGAGGACTACATGTCAGGCGGACAGCTTTGACTGACACA 400
1112 GGGTGGCAAGGCTTATCGCGGCTGAGAACTGAAGACACCTGGGGTCCACAGC 1171
401 GGAGAAAACCCGTTGTTTGACAGCCAGCTGGCTGTGATCAAAAATTCACAAATCA 460
1172 GGGGAAACCAATATGTGTGAGCAGAGGGCTGCMAAACCTTCCTCAACCCCTCG 1231
461 AACTTGAGAAACATTTGAAAGCGCAACATGAAATCAACAAACAAATATATGCACT 520
1232 GACCGGCAACACCAATGCGACCCAGCTCA--ACGAGAAACCTTACATCTGCAAG 1288
521 TTGGAAGACTGTGAAGAGA 539
1289 ATCCAGGCTGCACCAAGA 1307

RESULT 12
US-08-946-241B-1
Sequence 1, Application US/08946241B
Patent No. 5928941
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: MCA Nully, Megan M.
TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSER: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fastseed Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,241B
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creason, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 434...1843
OS-08-946-241B-1

```

```
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 434...1843
US-09-309-053-1

Query Match
Best Local Similarity 55.0%; Score 50.4; DB 3; Length 1889;
Matches 99: Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 235 CATCTGCTCCTTCCTGCTGACGCGCCCAATTACAGCAAGCCTGGAAGCTTGACGGCA 294
    || || || || || || || || || || || || || || || || || || || || ||
DB 1594 CACTTGCTGATTACGGGGCTGCGCAAAACCTACACAAAGAGTTCCCATCTCAGGACCA 1653
QY 295 CCTGTGCAAGCACACGGGGGAGAGACCATTTGTTGTGACTATGAAGGGTGTGCAAGGC 354
    || || || || || || || || || || || || || || || || || || || || ||
DB 1654 CCTGGCAACCCACACAGGTGAGAAACCTTACACACTGTGAGGACGCTGTGATGGA 1713
QY 355 CTTTCATCAGGACTACATCTGACCGCCGACATTTGACTACACACAGGAGAAAGCCGTT 414
    || || || || || || || || || || || || || || || || || || || || ||
DB 1714 ATTCCGCCCTCAGATGAACTGACACGAGCACTACCGTAACACACAGGAGCCGCCGTT 1773
```

RESULT 15

```
US-09-309-053-8
Sequence 8, Application US/09309053
Patent No. 6077933
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: MCA/Nulley, Megan M.
TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/309,053
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/946,241
FILING DATE: 07-OCT-1997
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creason, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
```

```
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 407...1843
US-09-309-053-8

Query Match
Best Local Similarity 55.0%; Score 50.4; DB 3; Length 1889;
Matches 99: Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 235 CATCTGCTCCTTCCTGCTGACGCGCCCAATTACAGCAAGCCTGGAAGCTTGACGGCA 294
    || || || || || || || || || || || || || || || || || || || || ||
DB 1594 CACTTGCTGATTACGGGGCTGCGCAAAACCTACACAAAGAGTTCCCATCTCAGGACCA 1653
QY 295 CCTGTGCAAGCACACGGGGGAGAGACCATTTGTTGTGACTATGAAGGGTGTGCAAGGC 354
    || || || || || || || || || || || || || || || || || || || || ||
DB 1654 CCTGGCAACCCACACAGGTGAGAAACCTTACACACTGTGAGGACGCTGTGATGGA 1713
QY 355 CTTTCATCAGGACTACATCTGACCGCCGACATTTGACTACACACAGGAGAAAGCCGTT 414
    || || || || || || || || || || || || || || || || || || || || ||
DB 1714 ATTCCGCCCTCAGATGAACTGACACGAGCACTACCGTAACACACAGGAGCCGCCGTT 1773
```

Search completed: February 10, 2003, 12:48:08
Job time : 109.257 secs

259 ACATTTCGCAATGAACCAAGTCTCCTAGGCATGCTGTTGTACATGATCCTGACAGAAG 200

Best Local Similarity 99.5%: Pred. No. 3 7e-103;

Matches 418: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 794 AACGCAAGATTACCTTAAGCAACACATGAAACATCCAGAAAGGATGTATGT 853
      |||
Db 1 AAACGGAAGATTACTTACCTTAAGCAACACATGAAACATCCAGAAAGGATGTATGT 60
      |||
OY 854 CGGTGCGAAGAGGCTGTGGAAGAACCTATCTGTTTAACTCCAAAGCCAT 913
      |||
Db 61 CGGTGCGAAGAGGCTGTGGAAGAACCTATCTGTTTAACTCCAAAGCCAT 120
      |||
OY 914 ATCTCTCTCTCCATGAGAAAGCCGCTTTTGTGTGAGACATCGTGCCTGACAA 973
      |||
Db 121 ATCTCTCTCTCCATGAGAAAGCCGCTTTTGTGTGAGACATCGTGCCTGACAA 180
      |||
OY 974 ACATTGCAATGAACAAAGTCTCAGTACGATCGTGTACATGATCCTGACAAAG 1033
      |||
Db 181 ACATTGCAATGAACAAAGTCTCAGTACGATCGTGTACATGATCCTGACAAAG 240
      |||
OY 1034 AAAATGAGCTCAAAAGTCAAAATCTGTGAAAAACGAGTTGGCTCTCATCTCACT 1093
      |||
Db 241 AAAATGAGCTCAAAAGTCAAAATCTGTGAAAAACGAGTTGGCTCTCATCTCACT 300
      |||
OY 1094 GGATATATCCCTCCCAAGAAAGCAAGGGCAAGGCTTATCTTGTGCAAAACGAGAG 1153
      |||
Db 301 GGATATATCCCTCCCAAGAAAGCAAGGGCAAGGCTTATCTTGTGCAAAACGAGAG 360
      |||
OY 1154 TCACCAACTGTGTGGAAGACAGATGCTCTGACAGTTCAGTACTTACCTTGACTAA 1213
      |||
Db 361 TCACCAACTGTGTGGAAGACAGATGCTCTGACAGTTCAGTACTTACCTTGACTAA 420
      |||
```

RESULT 5
US-09-796-692-4064

; Sequence 4064, Application US/09796692
; Publication No. US20020198362A1

; GENERAL INFORMATION:

; APPLICANT: Galger, Alexander
; APPLICANT: Algate, Paul A.

; APPLICANT: Mannion, Jane

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

; FILE REFERENCE: 2077.001200

; CURRENT APPLICATION NUMBER: US/09/796,692

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 60/186,126

; PRIOR FILING DATE: 2000-03-01

; PRIOR APPLICATION NUMBER: 60/190,479

; PRIOR FILING DATE: 2000-03-17

; PRIOR APPLICATION NUMBER: 60/200,545

; PRIOR FILING DATE: 2000-04-27

; PRIOR APPLICATION NUMBER: 60/200,303

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,779

; PRIOR FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/200,999

; PRIOR FILING DATE: 2000-05-01

; PRIOR APPLICATION NUMBER: 60/202,084

; PRIOR FILING DATE: 2000-05-04

; PRIOR APPLICATION NUMBER: 60/206,201

; PRIOR FILING DATE: 2000-05-22

; PRIOR APPLICATION NUMBER: 60/218,950

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: 60/222,903

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: 60/223,416

; PRIOR FILING DATE: 2000-08-04

; PRIOR APPLICATION NUMBER: 60/223,378

; PRIOR FILING DATE: 2000-08-07

; NUMBER OF SEQ ID NOS: 9597

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4064

; LENGTH: 234

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4064

Query Match 19.3%: Score 234; DB 9; Length 234;

Best Local Similarity 100.0%: Pred. No. 7.1e-54;

Matches 234: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 560 CTGAAATTCATCATGTCGACATACCAATGAACCTCTATTCAGTGTACCCAGAGA 619
      |||
Db 1 CTGAAATTCATCATGTCGACATACCAATGAACCTCTATTCAGTGTACCCAGAGA 60
      |||
OY 620 TGTGGGAAACACTTTGATCAACCCAGCAAGCTGAAAGCATGCCAAGGCCACGAGGGC 679
      |||
Db 61 TGTGGGAAACACTTTGATCAACCCAGCAAGCTGAAAGCATGCCAAGGCCACGAGGGC 120
      |||
OY 680 TATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATGAGCAACTTGTGAACAT 739
      |||
Db 121 TATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATGAGCAACTTGTGAACAT 180
      |||
OY 740 GTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATGCCGAAACATTT 793
      |||
Db 181 GTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATGCCGAAACATTT 234
      |||
```

RESULT 6
US-09-920-300A-610

; Sequence 610, Application US/09920300A
; Patent No. US2002016728A1

; GENERAL INFORMATION:

; APPLICANT: King, Gordon E.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Xu, Jiangchun

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.547

; CURRENT APPLICATION NUMBER: US/09/920,300A

; CURRENT FILING DATE: 2001-07-31

; NUMBER OF SEQ ID NOS: 1789

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 610

; LENGTH: 234

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-920-300A-610

Query Match 19.3%: Score 234; DB 10; Length 234;

Best Local Similarity 100.0%: Pred. No. 7.1e-54;

Matches 234: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 560 CTGAAATTCATCATGTCGACATACCAATGAACCTCTATTCAGTGTACCCAGAGA 619
      |||
Db 1 CTGAAATTCATCATGTCGACATACCAATGAACCTCTATTCAGTGTACCCAGAGA 60
      |||
OY 620 TGTGGGAAACACTTTGATCAACCCAGCAAGCTGAAAGCATGCCAAGGCCACGAGGGC 679
      |||
Db 61 TGTGGGAAACACTTTGATCAACCCAGCAAGCTGAAAGCATGCCAAGGCCACGAGGGC 120
      |||
OY 680 TATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATGAGCAACTTGTGAACAT 739
      |||
Db 121 TATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATGAGCAACTTGTGAACAT 180
      |||
OY 740 GTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATGCCGAAACATTT 793
      |||
Db 181 GTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATGCCGAAACATTT 234
      |||
```

RESULT 7
US-10-033-528-610

; Sequence 610, Application US/10033528
; Patent No. US20020131971A1

; GENERAL INFORMATION:

```

: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jlangchun
: APPLICANT: Secretist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547C1
: CURRENT APPLICATION NUMBER: US/10/033,528
: CURRENT FILING DATE: 2001-12-26
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 610
: LENGTH: 234
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-10-033-528-610

```

```

Query Match      19.38; Score 234; DB 12; Length 234;
Best Local Similarity 100.0%; Pred. No. 7,1e-54;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 560 CTGAATAATCCATGAGTCCAGCATACCAATGAACCTTATTCAGTGTACCCAGAGAGA 619
    |||||||
DB 1 CTGAATAATCCATGAGTCCAGCATACCAATGAACCTTATTCAGTGTACCCAGAGAGA 60
OY 620 TGTGGGAACACTTTCATCACCAGCAAGCTGAAGACATGCGCAGCCAGAGGCC 679
    |||||||
DB 61 TGTGGGAACACTTTCATCACCAGCAAGCTGAAGACATGCGCAGCCAGAGGCC 120
OY 680 TATGTATCTCAAAAAGAGATGTTCTTGTGGCAAAAACATGAGACGGAATCTGTGAACAT 739
    |||||||
DB 121 TATGTATCTCAAAAAGAGATGTTCTTGTGGCAAAAACATGAGACGGAATCTGTGAACAT 180
OY 740 GTGAGAGAACCCATAAGAGGAATACTGTGCAAGTGTGCCGGAACATTT 793
    |||||||
DB 181 GTGAGAGAACCCATAAGAGGAATACTGTGCAAGTGTGCCGGAACATTT 234

```

```

RESULT 8
US-09-292-758-68/c
: Sequence 68, Application US/09292758
: Publication No. US20020197602A1
: GENERAL INFORMATION:
: APPLICANT: Burner, Glenn C.
: APPLICANT: Brown, Joseph P.
: APPLICANT: Lifespan Biosciences, Inc.
: TITLE OF INVENTION: Nucleic Acid Sequences and Proteins
: FILE REFERENCE: 017473-001110US
: CURRENT APPLICATION NUMBER: US/09/292,758
: CURRENT FILING DATE: 1999-04-14
: EARLIER APPLICATION NUMBER: US 60/081,887
: EARLIER FILING DATE: 1998-04-15
: NUMBER OF SEQ ID NOS: 147
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 68
: LENGTH: 362
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-292-758-68

```

```

Query Match      18.6%; Score 226; DB 9; Length 362;
Best Local Similarity 96.7%; Pred. No. 1,3e-51;
Matches 263; Conservative 0; Mismatches 5; Indels 4; Gaps 3;

```

```

OY 946 TGTGTGACATGCTGGCTGTGGCAAAA--CATTTGCAATGAAACAAAGTGTGAC-TAG 1002
    |||||||
DB 362 TGTGTGACATGCTGGCTGTGGCAAAAACATTTGCAATGAAACAAAGTGTGAC-TAG 303
OY 1003 GCATCTCTTGTACATGATCCTGACAGAAAGAAATGAAGCTC-AAAGTCAAAAATCTGC 1061
    |||||||
DB 302 GCATCTCTTGTACATGATCCTGACAGAAAGAAATGAAGCTCAAAAATGAAGTCAAAAATCTGC 243

```

```

OY 1062 GTGAAAAACGAGTTTGCGCTCTCATCTCAGTGGATATATCCCTCCCAAAAGAAACAG 1121
    |||||||
DB 242 GTGAAAAACGAGTTTGCGCTCTCATCTCAGTGGATATATCCCTCCCAAAAGAAACAG 183
OY 1122 GCGAAGCTTATCTTTTGTGCAAAACGAGAGTCCACCAAGCTGTGGAAGCAAGATGC 1181
    |||||||
DB 182 GCGAAGCTTATCTTTTGTGCAAAACGAGAGTCCACCAAGCTGTGGAAGCAAGATGC 123
OY 1182 TCTGCACAGTTGCACTACTTACCTTGCTTA 1213
    |||||||
DB 122 TCTGCACAGTTGCACTACTTACCTTGCTTA 91

```

```

RESULT 9
US-09-995-973-2
: Sequence 2, Application US/09995973
: Publication No. US20030024006A1
: GENERAL INFORMATION:
: APPLICANT: CHOO, Yen
: APPLICANT: ULLMAN, Christopher G.
: TITLE OF INVENTION: GENE SWITCHES
: FILE REFERENCE: 8325-2003 / 97-US1
: CURRENT APPLICATION NUMBER: US/09/995,973
: CURRENT FILING DATE: 2002-03-19
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: LENGTH: 947
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: rFliiIA/zIf-VP64
: US-09-995-973-2

```

```

Query Match      12.6%; Score 152.4; DB 9; Length 947;
Best Local Similarity 57.6%; Pred. No. 1,9e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```

```

OY 208 GCCGCGCCCGCGCTCCACAGAGCTTCATCTGCTCTTCCCTGAGTCAGCGCCCAATTA 267
    |||||
DB 26 GCCGCTGGCGGTGTGATATAGCGGTACATCTGCTCTTCCCGCAGCTGCGCGCTTA 85
OY 268 CAGCAAGCTTGGAACTTTCAGCGGACCTGTGCAACACACAGCGGGAGAGACCAATTTGT 327
    |||||
DB 86 TAACAAGAACTGGAACTGCAAGCGCATCTGTCAAAACACACAGGAGAGAAACCAATTTGC 145
OY 328 TTGTACTATGAAGGTTGTGCAAGGCTTCATCAGGACTACATCTGAGCGGCACAT 387
    |||||
DB 146 ATGTAAAGAAAGAGATGTGAAAGGCTTTACTCGCTTCATCACTTAACCGGCACATC 205
OY 388 TGTGACTCACACAGGAAAGACCGCTTTGTTGTGCAAGCACTGCTGTGATCAAAATTT 447
    |||||
DB 206 ACTGACTCATCTACTGCGGAGAAACCTTCACATGTGACTGCGATGATGTGACTTGAGATT 265
OY 448 CACACAAATCAAACTTGAAGAAACATTTGAAGCAAAACATGAAATCAACAAACAA 507
    |||||
DB 266 TACTACAAAGCAAAACATGAAAGACCTTTAAGATTCATCAATCAACATGATGTGCGT 325
OY 508 ATATATATGACGTTTGAAGACTGTAAAGACCTTTAAGAAACATCAGCACTGAAAT 567
    |||||
DB 326 CTATGTGTCATTTTGGAGCTGTGGCAAGATTCGAAAGCAACAAATCAATTAAGGT 385
OY 568 CCATGATGCGAGCATACCAATGAACCTTAATTCAGGTGTACCAAGAAAGATGTGGGAA 627
    |||||
DB 386 TCATCAGTTGATGACACACAGAGCTGTGCGGTATGCTTGCGCTGTGAGTCTGCGATCG 445
OY 628 ACATTTGATCACCAGCAAGCTGAAGAAACATGCCAAGGCGCACAGAGGCTA 681
    |||||
DB 446 CCGCTTTCTCGCTCGATGAGCTTACCCTGATATCCGATATCCGATATCCGATATCCGAT 499

```

```

RESULT 10
US-09-732-348-5

```

```
; Sequence 5, Application US/09732348
; Patent No. US20020046419A1
; GENERAL INFORMATION:
; APPLICANT: Yen Choo, et al.
; TITLE OF INVENTION: Regulated Gene Expression in Plants
; FILE REFERENCE: 674538-2001
; CURRENT APPLICATION NUMBER: US/09/732,348
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 947
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (723)..(908)
; OTHER INFORMATION: transactivation domain of VP64, other features except c-myc tag
; NAME/KEY: misc.feature
; LOCATION: (909)..(938)
; OTHER INFORMATION: c-myc tag, other features except transactivation domain VP64 (115
US-09-732-348-5

Query Match      12.6%; Score 152.4; DB 10; Length 947;
Best Local Similarity 57.6%; Pred. No. 1.9e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 208 GCCGCGCCCGCGCTTCCACAGAGGTTTCATCTGCTCCCTCCCTGACGCGCCCAATTA 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26 GGGCGTGGCGGGTGTATTAAGCGGTACATCTGCTTTGCCGACGCGCGCTCTTA 85
QY 268 CAGCAAGCCCTGGAAGCTTGACGCGCACCTGTGCAACGACGCGGGAGAGACATTTGT 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 TACACAAAGCTGGAAGCTGACGCGCATCTGTGCMAACACAGGAGAGAACCATTTCC 145
QY 328 TTGTGACTGTGAGAGGTGTGCGACGCGCTTCATCAGGAGTACATCTGAGCGGCACAT 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 ATGTAGAGAAAGAGATGTGAGAAAGGCTTTACCTGCTTCATCACTTAACCGCCACATC 205
QY 388 TCTGACTCAGACAGAGAAAGCCGTTTGTGTGCGACGCACTGCTGTGATCAAAATTT 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 ACTGACTCTACTGTGGCGGAAAGAACTTCACATGTGACTCGGATGGATGTGACTTGAGATT 265
QY 448 CAACACAAGTAACCTTGAAGAAACATTTTGAACGCAACATGAAATCAACAAACAA 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266 TACTACAAAGGAAACATGAGAGCACTTTAACAGATTCATTAACATCAAGATCTGGCT 325
QY 508 ATATATATGACATTTTGAAGACTGTAGAGACTTTTAAGAAACATCAGAGCTGAAT 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326 CTATGTGTCCATTTTGGAACTGTGGCAAGCATTTCAAGAAACAAATCAATTAAGCT 385
QY 568 CCATCAGTCCGACATACCAATGAACCTCTATTCACTTACCCAGAGAGAGATGTGGAA 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 386 TCATCAGTTCACTCACAACAGAGCTGCGTATGCTTGGCCGTGTGAGATCGTGGAGATG 445
QY 628 ACACCTTGATCACCACGACATGAAACGACATGCCAAGGCCACGAGAGGCTA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 CCGCTTTTCTCGCTGATGAGCTTACCGGCAATATCCGATCCACACAGGCCA 499

RESULT 11
US-09-995-973-1
; Sequence 1, Application US/09995973
; Publication No. US20030024006A1
; GENERAL INFORMATION:
; APPLICANT: CHOO, Yen
; APPLICANT: ULLMAN, Christopher G.
; TITLE OF INVENTION: GENE SWITCHES
; FILE REFERENCE: 8325-2003 / G7-US1
; CURRENT APPLICATION NUMBER: US/09/995,973
; CURRENT FILING DATE: 2002-03-19
```

```
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 995
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
US-09-995-973-1

Query Match      12.6%; Score 152.4; DB 9; Length 995;
Best Local Similarity 57.6%; Pred. No. 1.9e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 208 GCCGCGCCCGCGCTTCCACAGAGGTTTCATCTGCTCCCTCCCTGACGCGCCCAATTA 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26 GGGCGTGGCGGGTGTATTAAGCGGTACATCTGCTTTGCCGACGCGCGCTCTTA 85
QY 268 CAGCAAGCCCTGGAAGCTTGACGCGCACCTGTGCAACGACGCGGGAGAGACCAATTTGT 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 TACACAAAGCTGGAAGCTGACGCGCATCTGTGCMAACACAGGAGAGAACCATTTCC 145
QY 328 TTGTGACTGTGAGAGGTGTGCGACGCGCTTCATCAGGAGTACATCTGAGCGGCACAT 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 ATGTAGAGAAAGAGATGTGAGAAAGGCTTTACCTGCTTCATCACTTAACCGCCACATC 205
QY 388 TCTGACTCAGACAGAGAAAGCCGTTTGTGTGCGACGCACTGCTGTGATCAAAATTT 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 ACTGACTCTACTGTGGCGGAAAGAACTTCACATGTGACTCGGATGGATGTGACTTGAGATT 265
QY 448 CAACACAAGTAACCTTGAAGAAACATTTTGAACGCAACATGAAATCAACAAACAA 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266 TACTACAAAGGCAACATGAGAGCACTTTTAACAGATTCATTAACATCAAGATCTGGCT 325
QY 508 ATATATATGACATTTTGAAGACTGTAGAGACTTTTAAGAAACATCAGAGCTGAAT 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326 CTATGTGTCCATTTTGGAACTGTGGCAAGCATTTCAAGAAACAAATCAATTAAGCT 385
QY 568 CCATCAGTCCGACATACCAATGAACCTCTATTCACTTACCCAGAGAGATGTGGAA 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 386 TCATCAGTTCACTCACAACAGAGCTGCGTATGCTTGGCCGTGTGAGATCGTGGAGATG 445
QY 628 ACACCTTGATCACCACGACATGAAACGACATGCCAAGGCCACGAGAGGCTA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 CCGCTTTTCTCGCTGATGAGCTTACCGGCAATATCCGATCCACACAGGCCA 499

RESULT 12
US-09-732-348-4
; Sequence 4, Application US/09732348
; Patent No. US20020046419A1
; GENERAL INFORMATION:
; APPLICANT: Yen Choo, et al.
; TITLE OF INVENTION: Regulated Gene Expression in Plants
; FILE REFERENCE: 674538-2001
; CURRENT APPLICATION NUMBER: US/09/732,348
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 995
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (15)..(17)
; OTHER INFORMATION: translational initiating ATG
; NAME/KEY: misc.feature
; LOCATION: (16)..(416)
; OTHER INFORMATION: Fingers 1 to 4 of TFIID
; NAME/KEY: misc.feature
; LOCATION: (308)..(416)
```

```

OTHER INFORMATION: spacer
NAME/KEY: misc.feature
LOCATION: (417)..(689)
OTHER INFORMATION: three fingers of zinc fingers protein ZNF268
NAME/KEY: misc.feature
LOCATION: (701)..(722)
OTHER INFORMATION: Nuclear Localization Signal
NAME/KEY: misc.feature
LOCATION: (957)..(986)
OTHER INFORMATION: c-myc tag
US-09-732-348-4

```

```

Query Match          12.6%; Score 152.4; DB 10; Length 995;
Best Local Similarity 57.6%; Pred. No. 1.9e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```

```

QY 208 GCGGGGGGGGGGGTCCAGAGAGTTCATCTCTCCCTCCGAGTGGAGGCCAATTA 267
DB 26 GCGGCTGCGGGTGTATTAAGCGGTACATCTCTCTCCGAGTGGAGGCCAATTA 85
QY 268 CAGCAAGCCTGGAGAGCTTGAACGGCAGCTGTGCAAGCAACGCGGGAGAGACATTTGT 327
DB 86 TAACAAGAACTGGAAAGCTGAGCGGCATCTGTGCAAAACACAGAGAAACCATTTCC 145
QY 328 TTGTACATATAAGGCTGTGGCAGGCTTCATCAGGAGTACCATCTGAGCCGCACAT 387
DB 146 ATGTAGAGAGAGAGATGTGAAAGGCTTACCTGCTTCATCAGTAAACCCGCATC 205
QY 388 TGTGACGACACAGAGAGAAAGCCGTTTGTGTGCAAGCAGCTGCTGTGATCAAAAT 447
DB 206 ACTCACTACTACTGGCGAGAAAGCTTCACATGTGAGTGGAGTGTGACTTGAGATT 265
QY 448 CAGCAAAATCAAACTTTGAAGAAACATTTGAAGCAAAATGAAGAAATCAACAAACAA 507
DB 266 TACTACAAAGCAAGATGAAGAAAGCACTTAAACAGATTCATTAACATATCATCTGCGT 325
QY 508 ATATATATGACGTTTGAAGAGCTGTGAAGACCTTTAAGAAACATCAGCAGTAAAT 567
DB 326 CTATGTGTGCTTTTGAAGACTGTGGCAAGATTCAGAAACAAATCAATTAAGGT 385
QY 568 CCATGAGTGCAGCATATCAATGAACCTTATTCAGTGTACCCAGAGAGATGTGGGA 627
DB 386 TCATCATGTTCACTGACACACAGACAGCTGCCGATGCTGCGCTGTGAGATCTCCGATCG 445
QY 628 ACAGTTTGCATCACCACAGAGCTGAAGCAATGCCAGAGCCGAGAGGCTTA 681
DB 446 CCGCTTTTCTGCTGATGAGCTTACCGCGCATATCCGATCCACAGAGGCA 499

```

RESULT 13

```

US-09-783-590-5443
Sequence 5443, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haselaine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2CI
CURRENT APPLICATION NUMBER: US/09/783,590
PRIOR FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
NUMBER OF SEQ ID NOS: 12465
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5443
LENGTH: 449
TYPE: DNA
ORGANISM: Homo sapiens

```

```

FEATURE:
NAME/KEY: misc.feature
LOCATION: (8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (53)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (66)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (83)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (181)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (245)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (312)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (359)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (364)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (366)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (397)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (407)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (448)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5443

```

```

Query Match          11.7%; Score 141.6; DB 10; Length 449;
Best Local Similarity 79.9%; Pred. No. 1e-28;
Matches 298; Conservative 0; Mismatches 58; Indels 17; Gaps 11;

```

```

QY 418 TTGTGAGGCACTGGC-TGTGATCAAAATTTCAACCAAAATCAAACTGAAGAAACATT 476
DB 63 TTGNCAGGCAATGGCTTGANATCAAAATTTCAACCAAAATCAAACTGAAGAAACATT 122
QY 477 TTG-TAAGCAAAATGAATCAACAAATATATATG-CAGTTTGAAGACTGTA 534
DB 123 TTGAAGCGCAAAATGAATCAACAAATATATATGCGAGTTTGAAGACTGTA 182
QY 535 -GAAGACCTTTAAGAA-----CATCAGAGCTGAATATCATCACTGAGCAGCATACCA 588
DB 183 GGAAGACCTTTAAGAAATCATCAAGAGCTGGAATATCATCACTGAGCAGCATACCA 242
QY 589 TG-AACCTTATT-CAAGTGTACCA-GGAAGATGTGGAAACACTTTG--CATCACC 643
DB 243 TGNAACTCTATTTCAGATGTACCCAGGAGGATGTGGAAACACTTTGCGATTCACC 302
QY 644 ACGAGCTGAAT--CGACATGCCAAGGCCACGAGGCTATGTATG--TCAAAAGGATGT 700
DB 303 ACGAGCTGNAAGAGACATGCCAAGGCCACGAGGCTATGTATGTTCAAAAAGGNTG 362
QY 701 TCGTTTGTGGCAAAACATGACGGAACCTGTGAACATGTGAGCAAAACCATTAAGAG 760
DB 363 TNCNTTGTGGCAAAACATGAGGAGGAGCTTGTGNACTGTGAGNAGAACCTTTAAGGG 422
QY 761 GAAATCTATGTG 773
DB 423 GATTATCTGTGAG 435

```

```

Query Match      6.7%; Score 81; DB 10; Length 368;
Best Local Similarity 98.8%; Pred. No. 2,2e-12;
Matches 81; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY 677 GCGCATGTATGTCMAAAGAGATGTTCCCTTTGTGCCCAAAACATGSGACGGAACCTTCTGAA 736
      |||
Db 356 GCGCATGTATGTCMAAAGAGATGTTCCCTTTGTGCCCAAAACATGSGAGGAACCTTCTGAA 297
      |||

QY 737 CATGTGAGAGAAACCCATTAAG 758
      |||
Db 296 CATGTGAGAGAAACCCATTAAG 275
      |||

Search completed: February 10, 2003, 12:58:42
Job time : 74.895 secs

```

```
Search completed: February 10, 2003, 12:58:42
Job time : 74.895 secs
```

```
Search completed: February 10, 2003, 12:58:42
Job time : 74.895 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:48:16 : Search time 3221.82 Seconds

(without alignments)
10957.076 Million cell updates/sec

Title: US-09-831-426c-4

Perfect score: 1213

Sequence: 1 gtccgcgcgcgcgcgcgaag.....cagctactacccttgcctaa 1213

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenDbml:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vtc:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1171.6	96.6	1381	9	HMUTITA
2	1170	96.5	1399	6	E10962
3	1088.8	89.8	1173	9	HSU20272
4	919.2	75.8	1098	9	HSU14134
5	889.8	73.4	1920	9	AK057993
6	719.2	59.3	1298	10	AF391799
7	691	57.0	1071	10	AF391798
8	560.6	46.2	996	10	BC032292
9	470.2	38.8	97979	9	AC004739
10	470.2	38.8	193126	9	AC006045
11	452.6	37.3	111722	9	AC074390
12	452.6	37.3	206622	2	AC016444
13	447	36.9	199517	9	AL353741
14	418.4	34.5	439	6	AX396111
15	416.8	34.4	439	6	AX396474
16	385.4	31.8	163421	2	AC012664
17	385.4	31.8	197709	9	AC073641
18	375.4	30.8	200885	2	AC012460
19	372.4	30.7	1693	9	AF265440
20	361.4	29.8	1518	5	XELFFI11A
21	355	29.3	1377	5	XBRF3A
22	350.8	28.9	1331	5	XELFINAB
23	340.6	28.1	1297	5	RPRFFI11A
24	340.6	28.1	1305	5	RANFFI11A
25	318.2	26.3	123905	9	AC116311
26	318.2	26.3	157165	2	AC104112
27	319.2	26.3	171007	2	AC021381
28	319.2	26.3	171705	2	AC113377
29	314.4	25.9	161441	9	AL137059
30	284.6	23.5	1314	5	BAPFI11A
31	278.6	23.0	67190	2	AC117566
32	234	19.3	234	6	AX396395
33	221.6	18.3	1129	5	IPOOF3A
34	202	16.7	162774	2	AC121381
35	202	16.7	182326	2	AC125693
36	197.6	16.3	112902	9	AC005230
37	177	14.6	272	6	AX324246
38	163	13.4	167342	2	AC124828
39	163	13.4	217225	10	AL513345
40	152.4	12.6	660	6	AX202580
41	152.4	12.6	947	6	AX053174
42	152.4	12.6	947	6	AX059770
43	152.4	12.6	947	6	AX201763
44	152.4	12.6	947	6	AX201936
45	152.4	12.6	995	6	AX053173

ALIGNMENTS

RESULT 1

HMUTITA

LOCUS

DEFINITION Human GTF3A mRNA for Xenopus transcription factor ITTA homologue, complete cds.

ACCESSION D332257.1 GI:1000446

VERSION D332257

KEYWORDS GTF3A; Xenopus transcription factor ITTA homologue.

SOURCE Homo sapiens cDNA to mRNA, clone J1b11library of T.Fujiwara, S.Shin and Y.Nakamura clone:39H11.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1381)

AUTHORS Arakawa,H., Nagase,H., Hayashi,N., Ogawa,M., Nagata,M.,

TITLE	Fujiwara, T., Takahashi, E., Shlin, S. and Nakamura, Y.
JOURNAL	Molecular cloning, characterization, and chromosomal mapping of a novel human gene (GFP3a) that is highly homologous to Xenopus transcription factor TTA
MEDLINE	Cytogenet. Cell Genet. 70 (3-4), 225-228 (1995)
REFERENCE	95309028
AUTHORS	2 (bases 1 to 1381)
TITLE	Nakamura, Y.
JOURNAL	Direct Submission
REFERENCE	Submitted (22-JUL-1994) Yusuke Nakamura, Cancer Institute, Department of Biochemistry, 1-37-1 Kam-1-Ikebukuro, Toshima-ku, Tokyo 170, Japan (E-mail:nakamura@ganvil.jicr.or.jp, Tel:03-3918-0111(ex.4501), Fax:03-3918-0342)

Query Match	96.58;	Score 1170;	DB 6;	Length 1399;
Best Local Similarity	99.38;	Pred. No. 1.1e-252;		
Matches 1206; Conservative	0;	Mismatches 5;	Indels 3;	Gaps 3

OY	721	GATGACCTTCGAACAATGTGAGAGAACCATAAAGAGAAATCTATGTGAAGTATG	780
Db	799	GACGAGACTTCTBMACNTGTGAGAGAAAACCCATTAAAGAGAAATCTATGTGAAGTATG	858
OY	781	CCGAAACAATTTAAAGCAAAAGATTACCTTTAACCAACACATGAAAATCTCATGCCACGA	840
Db	859	CCCGAAACAATTTAAAGCMAAGATTACCTTAACCAACACATGAAAATCTCATGCCACGA	918
OY	841	AAGGATGTATGTGCCTGTCCAAAGAGAGCGTGGAAGAACCCTACTACTGTGTTAA	900
Db	919	AAGGATGTATGTGCCTGTCCAAAGAGAGCGTGGAAGAACCCTACTACTACTGTGTTAA	978
OY	901	TCTCCAAAGCCATATCTCTCTCTCCATGAGAGAAAGCCGCCCTTTTGTGTGTAACATGC	960
Db	979	TCTCCAAAGCCATATCTCTCTCTCCATGAGAGAAAGCCGCCCTTTTGTGTGTAACATGC	1038
OY	961	TGGCTGTGGCAAAACATTTTGCAATGAAACAAAGTCTCACTAGGATCTGTTGTATACATG	1020
Db	1039	TGGCTGTGGCAAAACATTTTGCAATGAAACAAAGTCTCACTAGGATCTGTTGTATACATG	1098
OY	1021	TCCGACAAACAATAAATGAACTGCAAGCTAAAAAATCTCTGTAAGAAAC-GGAGTTGG	1079
Db	1099	TCCGACAAACAATAAATGAACTGCAAGCTAAAAAATCTCTGTAAGAAAGGAGGTTGG	1158
OY	1080	CCTCTCATCTCAGTGTGATATATCCCTCCCAAAGGAAACAAGGCAAGCTTATCTTTGT	1139
Db	1159	CCTCTCATCTCAGTGTGATATAT-CCTCCCAAAGGAAACAAGGCAAGCTTATCTTTGT	1217
OY	1140	GTCMAAAGSAGAGTCAACCACACTGTGTGGAAGACAAAGATCTCTGCAGCTGGCAGTAC	1199
Db	1218	GTCMAAAGSAGAGTCAACCACACTGTGTGGAAGACAAAGATCTCTGCAGCTGGCAGTAC	1277
OY	1200	TTACCTTGGCTAA 1213 	
Db	1278	TTACCTTGGCTAA 1291 	
<hr/>			
RESULT_3			
LOCUS	HSU20272	1173 bp mRNA linear	PRI 26-SEP-1995
DEFINITION	Human DNA/RNA-binding protein mRNA, partial cds.		
ACCESSION	U020272		
VERSION	U020272.1 GI:644870		
KEYWORDS	.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (sites) Drew,P.D., Nagle,J.W., Canning,R.D., Ozato,K., Biddison,W.E. and Becker,K.G. Cloning and expression analysis of a human cDNA homologous to Xenopus TrfIIA JOURNAL MEDLINE Gene 159 (2), 215-218 (1995) PUBMED 95347600 7622052 2 (bases 1 to 1173) Becker,K.G. Direct Submssion Submitted (25-JAN-1995) Kevin G. Becker, Nat. Inst. of Neur. Diseases and Stroke/NIH, Neuroimmunology Branch, 9000 Rockville Pike, Bethesda, MD 20892, USA LOCATION/Qualifiers 1. .1173 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CH2-34.10" /sex="female" /tissue_type="brain; hippocampus" /clone_id="Stratagene #936205" /dev_stage="2 years old"		
FEATURES	source		
CDS	<1..1092		

```
Accession Number P03001"
/codon_start=1
/product="DNA/RNA-binding protein"
/protein_id="AA75623.1"
/db_xref="GI:644871"
/translation="PRAVVAESVSLITLADAFIAGESSAPLPPEPALPRFICFPD
CSAVYSAAMKLDALCKHTGERPFDVGEQCAIRIYHLSRIHLHTGEPKPVCA
NGCDOKFNTSNLKKHERRHNOOKOYICSEEDCKTFKHOOGLKHOOCHNEPLF
KCTOEQCKHFAAPSKLRKRAKHEGYCOOKCSFVATPTTELLKHVREHKEPICE
VCKRTFRADYLYKOHKHTHAPERDVCRCRCECGTITTYVNLDSHLSHRESRPV
CEHAGCGKTPAMKQSLTRHAYVDDPKMKMKKAKSKREKSLASHLSGYIPPRKOG
OGSLCKNGESPNCVEDKMLSTVAVLTLG"
misc_feature
118..184
/note="encodes zinc-finger"
misc_feature
208..274
/note="encodes zinc-finger"
misc_feature
298..367
/note="encodes zinc-finger"
misc_feature
394..460
/note="encodes zinc-finger"
misc_feature
484..550
/note="encodes zinc-finger"
misc_feature
565..631
/note="encodes zinc-finger"
misc_feature
649..709
/note="encodes zinc-finger"
misc_feature
736..805
/note="encodes zinc-finger"
misc_feature
829..895
/note="encodes zinc-finger"
misc_feature
363 a 283 c 265 g 262 t
BASE COUNT
ORIGIN
```

```
Query Match 89.8% Score 1088.8; DB 9; Length 1173;
Best Local Similarity 99.8%; Pred. No. 1.9e-234;
Matches 1090; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 122 CCGCGGCGCGTGTGTCGCCAGTGTGTCCTTGACATGCGCCGACCGCTTCATTGCA 181
DB 1 CCGCGGCGCGTGTGTCGCCAGTGTGTCCTTGACATGCGCCGACCGCTTCATTGCA 60
QY 182 GCGGCGGAGAGTGTGTCGCCAGTGTGTCCTTGACATGCGCCGACCGCTTCATTGCA 241
DB 61 GCGGCGGAGAGTGTGTCGCCAGTGTGTCCTTGACATGCGCCGACCGCTTCATTGCA 120
QY 242 TCCTTCCCTGACTGTGAGCGCCCAATTACAGCAAAAGCCTGGAAGCTTGACGACCTGTGC 301
DB 121 TCCTTCCCTGACTGTGAGCGCCCAATTACAGCAAAAGCCTGGAAGCTTGACGACCTGTGC 180
QY 302 AAGCACAGGCGGAGAGCAATTGTTGTGACATGANGGGTGTGCGCAAGGCGTTCATC 361
DB 181 AAGCACAGGCGGAGAGCAATTGTTGTGACATGANGGGTGTGCGCAAGGCGTTCATC 240
QY 362 ACGGACTACATCTGAGCGCGCCACATTTCTGACTACAGAGAGAAAAGCCCTTTGTTGT 421
DB 241 ACGGACTACATCTGAGCGCGCCACATTTCTGACTACAGAGAGAAAAGCCCTTTGTTGT 300
QY 422 GCAGCAGTGGCTGTGATCAAAATTCACACAAATTCAAACTGGAAGAAACATTTTGA 481
DB 301 GCAGCAGTGGCTGTGATCAAAATTCACACAAATTCAAACTGGAAGAAACATTTTGA 360
QY 482 CGCAAACTGAAAATCAACAAAACAAATATATATGAGTTTGAACACTGTAAAGAAC 541
DB 361 CGCAAACTGAAAATCAACAAAACAAATATATATGAGTTTGAACACTGTAAAGAAC 420
QY 542 TTATGAAGAAATCGAGAGCTGAATTCATGAGCAGATGACGAATGAACCTGATTC 601
DB 421 TTATGAAGAAATCGAGAGCTGAATTCATGAGCAGATGACGAATGAACCTGATTC 480
QY 602 AAGTGTACCCAGAGAGATGTGGAAAACACTTTGCATCAACCCAGAACCTGAAGACAT 661
DB 481 AAGTGTACCCAGAGAGATGTGGAAAACACTTTGCATCAACCCAGAACCTGAAGACAT 540
```

```
QY 652 GCCAAGCCCCACAGAGGCTATGTATGTCAAAAAGATGTTCCITTTGGCAAAACATGG 721
DB 541 GCCAAGCCCCACAGAGGCTATGTATGTCAAAAAGATGTTCCITTTGGCAAAACATGG 600
QY 722 ACGGAATCTTGAACATGTAGAGAAACCCATTAAGAGAAATTAATGTAGATATGC 781
DB 601 ACGGAATCTTGAACATGTAGAGAAACCCATTAAGAGAAATTAATGTAGATATGC 660
QY 782 CGGAAAACATTTAAACGGAAGATTACCTTAAGCAACACTGAAAACATATGCCAGAA 841
DB 661 CGGAAAACATTTAAACGGAAGATTACCTTAAGCAACACTGAAAACATATGCCAGAA 720
QY 842 ACGATGTATGTGCTGTGCCAAGAGAGGCTGTGGAAGAACTATCTGTGTTAT 901
DB 721 ACGATGTATGTGCTGTGCCAAGAGAGGCTGTGGAAGAACTATCTGTGTTAT 780
QY 902 CTCGAAGGCATATCTCTCTCTCATGAGAGAAAGCCGCTTTGTGTGTAACATGCT 961
DB 781 CTCGAAGGCATATCTCTCTCTCATGAGAGAAAGCCGCTTTGTGTGTAACATGCT 840
QY 962 GCCTGTGGCAAAACATTTGCAATGAACAAAGTCTCACTAGCATGCTGTGTACATGAT 1021
DB 841 GCCTGTGGCAAAACATTTGCAATGAACAAAGTCTCACTAGCATGCTGTGTACATGAT 900
QY 1022 CCTGACAGAGAAATGAATGAAGTCAAAAGTCTGTGTAAGAAAGAGGATTTGGCC 1081
DB 901 CCTGACAGAGAAATGAATGAAGTCAAAAGTCTGTGTAAGAAAGAGGATTTGGCC 960
QY 1082 TCTCATCTCAGTGTATATATCCCTCCCAAAAGAAACAAGGCAAGGCTTAATCTTGTGT 1141
DB 961 TCTCATCTCAGTGTATATATCCCTCCCAAAAGAAACAAGGCAAGGCTTAATCTTGTGT 1020
QY 1142 CAAACGGAGAGTCCCAACTGTGTGGAAGACAAAGTGTCTGCAACTGCATGACTACT 1201
DB 1021 CAAACGGAGAGTCCCAACTGTGTGGAAGACAAAGTGTCTGCAACTGCATGACTACT 1080
QY 1202 ACCCTTGGCTAA 1213
DB 1081 ACCCTTGGCTAA 1092
```

```
RESULT 4
LOCUS HS014134 1098 bp mRNA linear PRI 01-OCT-1994
DEFINITION Human transcription factor IIA (HTFIIA) mRNA, partial cds.
ACCESSION U14134
VERSION U14134.1 GI:551534
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1098)
REFERENCE
1 Drew,P.D., Nagle,J.W., Canning,R.D., Ozato,K., Biddison,W.E. and
Becker,K.G.
Cloning and expression analysis of a cDNA encoding human TFIIA
2 (bases 1 to 1098)
REFERENCE
Becker,K.G.
Direct Submission
Submitted (29-AUG-1994) Kevin G. Becker, National Institute of
Neurological Diseases and Stroke/NIH, Neuroimmunology Branch, 9000
Rockville Pike, Bethesda, MD 20892, USA
LOCATION/Qualifiers
1..1098
FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="C2H2-34"
/sex="2 yr old female"
/lisue_type="brain"
/clone_id="Human hippocampus cDNA library, Stratagene
Number 936205"
1..1098
```

gene

```

CDS
/gene="HtPflIIA"
<1..1017
/note="HtPflIIA"
/note="HtPflIIA: similar to Xenopus transcription factor
IIIA, Swiss-Prot Accession Number P03001"
/codon_start=1
/product="transcription factor IIIA"
/protein_id="AA021873.1"
/db_xref="GI:551535"
/translation="PPAVVAESYSSTLTADATIAGESAPPPPPALPRFETSPD
CSANYSKAMKLDHLCITGERPVCDDYEGCGKATFDHLSRHLLHTGKPRVCA
NGCQDNKTSNLKHEERKHEKNOQOYICSEDCRKTFRKHQDLKHOCHEPELE
KLCCQKGSVPAAKTMTELLKRVRETHEBELLCEVCRKTRKRDYLRKHMTNABEYV
CRCPREGCRPTVTVENLOSHLISFHEESRPVCEAGCGKTFPAKOSLTHRAVYHP
DKKKMLKVKRSREKSLASLSGYIPRKQOGQSLCQNESEPNVEDKMLSTAV
LTLC"
misc_feature
118..186
/note="HtPflIIA"
/note="encodes zinc-finger"
misc_feature
208..276
/note="HtPflIIA"
/note="encodes zinc-finger"
misc_feature
298..369
/note="HtPflIIA"
/note="encodes zinc-finger"
misc_feature
394..462
/note="HtPflIIA"
/note="encodes zinc-finger"
misc_feature
490..558
/note="HtPflIIA"
/note="encodes zinc-finger"
misc_feature
574..636
/note="HtPflIIA"
/note="encodes zinc-finger"
misc_feature
661..732
/note="HtPflIIA"
/note="encodes zinc-finger"
misc_feature
754..822
/note="HtPflIIA"
/note="encodes zinc-finger"
BASE COUNT 339 a 263 c 244 g 252 t
ORIGIN
Query Match 75.8%, Score 919.2, DB 9, Length 1098;
Best local Similarity 92.4%; Pred. No. 2.8e-196;
Matches 1009; Conservative 0; Mismatches 8; Indels 75; Gaps 1:
Oy 122 CGCGCGCGCGTGTGCGGAGTGTGTCCTTGACATCGCGCGAGCGTTGATTGCA 181
Db 1 CGCGCGCGCGTGTGCGGAGTGTGTCCTTGACATCGCGCGAGCGTTGATTGCA 60
Oy 182 GCGGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCGCTTCCAGAGCTTCATCTGC 241
Db 61 GCGGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCGCTTCCAGAGCTTCATCTGC 120
Oy 242 TCCCTTCCTGACTGAGCGGCAATPACGAAACCGTGAAGCTTGACGCGGACCTGTGC 301
Db 121 TCCCTTCCTGACTGAGCGGCAATPACGAAACCGTGAAGCTTGACGCGGACCTGTGC 180
Oy 302 AAGCACACGGGGAGAGACCATTTGTTGTGACTATGTAAGGGGTGTGGCAAGGCGCTTCATC 361
Db 181 AAGCACACGGGGAGAGACCATTTGTTGTGACTATGTAAGGGGTGTGGCAAGGCGCTTCATC 240
Oy 362 AGGACTACATGTGAGCGGCGACATCTGTGACTCAGACAGAGAAAGCCGTTGTTGT 421
Db 241 AGGACTACATGTGAGCGGCGACATCTGTGACTCAGACAGAGAAAGCCGTTGTTGT 300
Oy 422 GCGAGCAGTGGCTGTGATCAAAATTCACACAAATTCAACTTGAAGAAACATTTTGA 481
Db 301 GCGAGCAGTGGCTGTGATCAAAATTCACACAAATTCAACTTGAAGAAACATTTTGA 360
Oy 482 CGGAACATGAAATCAACAAACAAATATATGACAGTTTGAAGACTGTAGAGAGAC 541
Db 360 CGGAACATGAAATCAACAAACAAATATATGACAGTTTGAAGACTGTAGAGAGAC 420

```

```

Db 361 CGGAACATGAAATCAACAAACAAATATATGACAGTTTGAAGACTGTAGAGAGAC 420
Oy 542 TTTAAGAAACATCAGCAGCTGTAATAATCCATCAGTGCAGCATACGAATCAACCTGTATTC 601
Db 421 TTTAAGAAACATCAGCAGCTGTAATAATCCATCAGTGCAGCATACGAATCAACCTGTATTC 480
Oy 602 AAGTGTACCCAGAGAGATGTGGGAAACACTTTCATCACCAGCAAGGTGAAGAGACAT 661
Db 481 AAC-----CTATGTTGTCAAAAAGAGATGTCCTTTGTGCAAAAACATATG 483
Oy 662 GCCAAGCCCGACAGGCGTATGTATGTCAAAAAGATGTTCTTTGTGCAAAAACATATG 721
Db 484 -----CTATGTTGTCAAAAAGAGATGTCCTTTGTGCAAAAACATATG 525
Oy 722 AGGACTTCCTGAACATGTGAGAGAAACCCTAAGAGAAATATATGATGAGATATGC 781
Db 526 AGGAACTTCCTGAACATGTGAGAGAAACCCTAAGAGAAATATATGATGAGATATGC 585
Oy 782 CGGAACATTTTAAAGCAAGATTACCTTAAGCAACATGAAATCATGCCAGAA 841
Db 586 CGGAACATTTTAAAGCAAGATTACCTTAAGCAACATGAAATCATGCCAGAA 645
Oy 842 AGGATGTATGTGCGTGTCAAGAGAGGCTGTGAGAGAACTATCTGTTTAT 901
Db 646 AGGATGTATGTGCGTGTCAAGAGAGGCTGTGAGAGAACTATCTGTTTAT 705
Oy 902 CTCGAAAGCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGTAACATGCT 961
Db 706 CTCGAAAGCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGTAACATGCT 765
Oy 962 GGGTGTGGCAAAACATTTGCAATGAACAAAGTCTCACTGAGATCTGTGATATGAT 1021
Db 766 GGGTGTGGCAAAACATTTGCAATGAACAAAGTCTCACTGAGATCTGTGATATGAT 825
Oy 1022 CTTGACAGAGAGAAATGAAAGTCAAAAGTCAAAATCTGTGAAAAAGGAGTTTGGC 1081
Db 826 CTTGACAGAGAGAAATGAAAGTCAAAAGTCAAAATCTGTGAAAAAGGAGTTTGGC 885
Oy 1082 TCTCATCTCAGTGTATATCTCTCTCCCAAGAGAAACAGGAGAGCTTATCTTGTGT 1141
Db 886 TCTCATCTCAGTGTATATCTCTCTCCCAAGAGAAACAGGAGAGCTTATCTTGTGT 945
Oy 1142 CAAAGCGAGAGTCAACCACTGTGTGAGAGACAAGTGTGCAAGTTCAGTACTT 1201
Db 946 CAAAGCGAGAGTCAACCACTGTGTGAGAGACAAGTGTGCAAGTTCAGTACTT 1005
Oy 1202 ACCCTTGGCTAA 1213
Db 1006 ACCCTTGGCTAA 1017
RESULT 5
AK057993 1920 bp mRNA linear PRI 31-OCT-2001
LOCUS Homo sapiens cDNA FLJ25264 fis, clone STM05057, highly similar to
DEFINITION TRANSCRIPTION FACTOR IIIA.
ACCESSION AK057993.1 GI:16533986
VERSION AK057993.1
SOURCE Homo sapiens stomach mucosa cDNA to mRNA, clone_11b:STM
clone:STM05057.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidi; Homo.
REFERENCE 1
AUTHORS Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,
Ono,Y., Holuca,T., Hirooka,S., Murakawa,K., Takiguchi,S.,
Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
Yamashita,H., Chiba,Y., Suzuki,Y., Hata,H., Nakagawa,K., Mizuno,S.,
Morinaga,M., Kawamura,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,
Nishikawa,T., Sugiyama,A., Kawakami,B., Nagai,K., Isogai,T. and
Sugano,S.
TITLE NED0 human cDNA sequencing project

```

```

JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1920)
AUTHORS      Sugano, S. and Suzuki, Y.
TITLE        Direct Submission
JOURNAL      Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science,
              University of Tokyo, Laboratory of Genome Structure, Human Genome
              Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
              (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
              Fax:81-3-5449-5416)
COMMENT      NEDO human cDNA sequencing project supported by Ministry of
              Economy, Trade and Industry of Japan; cDNA full insert sequencing:
              Research Association for Biotechnology (RAB); cDNA library
              construction and 5'-end one pass sequencing: Institute of Medical
              Science, University of Tokyo, Laboratory of Genome Structure, Human
              Genome Center; 3'-end one pass sequencing: RAB; clone selection for
              full insert sequencing: RAB and Helix Research Institute.

FEATURES
  source
    1..1920
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="STM05057"
    /tissue_type="stomach mucosa"
    /clone_id="STM"
    /note="cloning vector: pME185FL3"

BASE COUNT   520 a 462 c 438 g 500 t
ORIGIN
Query Match 73.48; Score 889.8; DB 9; Length 1920;
Best Local Similarity 99.28; Pred. No. 1.2e-189;
Matches 894; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 313 GGAGAGACCATTTGTTGTGACTATGAGAGGTGTGGCAAGGCTTCATCAGAGCATACCA 372
    |||||||
DB 928 GCGTAGACCATTTGTTGTGACTATGAGAGGTGTGGCAAGGCTTCATCAGAGCATACCA 987
    |||||||
QY 373 TCTGAGCCGCCATCTTGACTCAGACAGAGAAAGCCGTTGTTGTCAGCCACTGG 432
    |||||||
DB 988 TGTGAGCCGCCATCTTGACTCAGACAGAGAAAGCCGTTGTTGTCAGCCACTGG 1047
    |||||||
QY 433 CTGTGATCAAAATTCAGACAAATCAATCAATCAATCAATTTGAAGCCAAACATGA 492
    |||||||
DB 1048 CTGTGATCAAAATTCAGACAAATCAATCAATCAATCAATTTGAAGCCAAACATGA 1107
    |||||||
QY 493 AATCAAGAAAAACAATATATATGCACTTTGAAGACTGTAGAAGACCTTTAAAGAA 552
    |||||||
DB 1108 AATCAAGAAAAACAATATATATGCACTTTGAAGACTGTAGAAGACCTTTAAAGAA 1167
    |||||||
QY 553 TCAGCAGCTGAAATTCATCATGCTCCAGCATACCAATGAACCTCTATTCAAGTATACCA 612
    |||||||
DB 1168 TCAGCAGCTGAAATTCATCATGCTCCAGCATACCAATGAACCTCTATTCAAGTATACCA 1227
    |||||||
QY 613 GGAAGAGTGGGAAACCTTTCATCACCAGCAGAGCTGAAACGACATGCCAAGGCCCA 672
    |||||||
DB 1228 GGAAGAGTGGGAAACCTTTCATCACCAGCAGAGCTGAAACGACATGCCAAGGCCCA 1287
    |||||||
QY 673 CGAGGCTATGATCTCAAAAAGAGATGTCCTTTGGCAAAAACATGAGCGAACTTCT 732
    |||||||
DB 1288 CGAGGCTATGATCTCAAAAAGAGATGTCCTTTGGCAAAAACATGAGCGAACTTCT 1347
    |||||||
QY 733 GAAACATGTGAGAAACCCATTAAGAGAGAAATATATATGTAAGTATGCCGGAACATTT 792
    |||||||
DB 1348 GAAACATGTGAGAAACCCATTAAGAGAGAAATATATATGTAAGTATGCCGGAACATTT 1407
    |||||||
QY 793 TAAAGCAAAAGTTACTTTAAGCAACATGAAAGCTATGCCCGGAAAGGATGATATG 852
    |||||||
DB 1408 TAAAGCAAAAGTTACTTTAAGCAACATGAAAGCTATGCCCGGAAAGGATGATATG 1467
    |||||||
QY 853 TGGCTGTCCAGAGAGAGCTGTGAAGAACCTATACCTACTGTTTAACTCTCAAAAGCA 912
    |||||||
DB 1468 TGGCTGTCCAGAGAGAGCTGTGAAGAACCTATACCTACTGTTTAACTCTCAAAAGCA 1527
    |||||||
QY 913 TATCCCTCTCTTCATGAGAAAGCCGCTTTGTGTGTGAACATGCTGGCTGGCAA 972
    |||||||

```

```

DB 1528 TATCCCTCTCTTCATGAGAAAGCCGCTTTGTGTGTGAACATGCTGGCTGGCAA 1587
    |||||||
QY 973 AACATTTGCATGAACAAAGCTCTCAGCATGCTGTTGTACATGATCTGCACAGAA 1032
    |||||||
DB 1588 AACATTTGCATGAACAAAGCTCTCAGCATGCTGTTGTACATGATCTGCACAGAA 1647
    |||||||
QY 1033 GAAATGAAGCTCAAGTCAAAATATCTCGTGAAGAAACGAGTTTGGCTCTCATCTCAG 1092
    |||||||
DB 1648 GAAATGAAGCTCAAGTCAAAATATCTCGTGAAGAAACGAGTTTGGCTCTCATCTCAG 1707
    |||||||
QY 1093 TCGATATATCCCTCCCAAGAGAAACAGAGGCAAGCTTATCTTTGTCAAAACGAGA 1152
    |||||||
DB 1708 TCGATATATCCCTCCCAAGAGAAACAGAGGCAAGCTTATCTTTGTCAAAACGAGA 1767
    |||||||
QY 1153 GTCACCAACTGTGTGAAGACAGATGCTCTGACAGTGTGACAGTACCTTGGGCTA 1212
    |||||||
DB 1768 GTCACCAACTGTGTGAAGACAGATGCTCTGACAGTGTGACAGTACCTTGGGCTA 1827
    |||||||
QY 1213 A 1213
DB 1828 A 1828

RESULT 6
AF391799 1298 bp mRNA linear ROD 31-JAN-2002
LOCUS AF391799
DEFINITION Mus musculus transcription factor I11a mRNA, partial cds.
ACCESSION AF391799
VERSION AF391799.1 GI:18448381
KEYWORDS
SOURCE
  Mus musculus.
  Mus musculus.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 1298)
  Hanas, J.S., Hocker, J.R., Cheng, Y.G., Lerner, M.R., Brackett, D.J.,
  Lightfoot, S.A., Hanas, R.J., Madhusudan, K.T., and Moreland, R.J.,
  cDNA cloning, DNA binding, and evolution of mammalian transcription
  factor I11a
  JOURNAL Gene 282 (1-2), 43-52 (2002)
  MEDLINE 21673987
  PUBMED 11814676
REFERENCE
  2 (bases 1 to 1298)
  Hanas, J.S., Hocker, J.R., Lerner, M.R., Brackett, D.J.,
  Lightfoot, S.A., Hanas, R.J., Madhusudan, K., and Moreland, R.J.,
  Direct Submission
  JOURNAL Submitted (14-JUN-2001) Biochemistry/Molecular Biology, University
    of Oklahoma Health Science Center, 800 Research Parkway, Suite 448,
    Oklahoma City, OK 73104, USA
FEATURES
  source
    1..1298
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /note="Tf11a: Cys2His2 zinc finger protein; similar to
    Xenopus laevis transcription factor I11a Swiss-Prot
    Accession Number P03001"
    /codon_start=1
    /product="transcription factor I11a"
    /protein_id="AAL69686.1"
    /translation="EFGTEVQSLRAVAVCPVSRGRHVAARQRPDRVALPRVSA
    EAVSLLTIADAFVAGCEGPAPRALPSPFICSPDCSASYNKAMKLDANLCKITGR
    PFVCDYEGCKAPFIRYHLSRHLITGSKPFVCDGDCNOKFPTKSLNKLKRRKG
    NPOKQILCYEGKKAFFKRRHDLRTHOCHTSEPLFRCTHGGCKHRSRLKRIE
    VHEGYLCQKCSFMGKTWTELLKHMEAKEDITCNQGRMPKRRDYLKQHKTHARE
    RDYRCPKRCQCGTTTTYFENLSHLSFHEKRPVCEHACGCTPAKQSLKHSYV
    HDPKRMRKLVRAPREKRSLSRSGTFPPKRPDPSLPNASESSSPENQLPP
    PAALLVC"
BASE COUNT 342 a 368 c 358 g 230 t
ORIGIN
Query Match 59.38; Score 719.2; DB 10; Length 1298;

```

Best Local Similarity 79.5%: Pred. No. 2.8e-151:	
Matches 879: Conservative 0: Mismatches 218: Indels 9: Gaps 2:	
QY 110	GGCGCCCTGATCCGCCGCCGCGTGGTCCGAGTCGCTTGCATCGCCGAC 169
Db 103	GTGCGCCGTGAGACCCGCGGTGTCACTCGCGGAAGCGGTGCTGCTGACATCGCGAT 162
QY 170	GGCTCATTCGACGGGGGAGAGCTCAGCTCCGACCCCGCGCGCGCGCTTCCGAG 229
Db 163	GGCTTCGTCGGGGCTGTGAGGGCCCGCGCG-----CGCGCGCGCGCGTCCGAC 216
QY 230	AGGTCATCTGCTCTCTCCCTGACTGACGGCCAAATTCAGCAAGCCCTGGAAGCTTGAC 289
Db 217	AGGTCATCTGCTCTCTCTCCGACTGACGGCCAGTTACAAACACCTTGGAAGCTAGAC 276
QY 290	GGGCACTGTGTGAACGAGGGGAGAGACCAATTCCTGTGTCATGATGAGGCTGGC 349
Db 277	GGGCACTGTGTGAACGAGGGGAGAGGCAATTCCTGTGTCATGATGAGGCTGGC 336
QY 350	AAGGCTTCATCAGGAGTACATCTGAGCCGCCCATCTCTGACTCAGACAGAGAAAG 409
Db 337	AAGGCTTCATCAGAGACTACCATCTGAGCCGCGCATGTCTGTTCACACCGGGGAAAG 396
QY 410	CGGTTGTTTGGGACGACAGTGGCTGTGATCAAAATTCAGCAAAATCAAACTTGAAG 469
Db 397	CGGTTGTTTGGGACAGATGATGCTGTATCAGAAATTCAGCAAAATCAAACTTGAAG 456
QY 470	AAACATTTTGAACGCAACATGAAATTCACAAAACATATATATGCACTTTTGAAGAC 529
Db 457	AAACATTTTGAACGCAACATGAAATTCACAAAACATATATATGCACTTTTGAAGAC 516
QY 530	TGTAAGAAGACCTTTAAGAAACATCAGAGCTGAAATTCATAGGTCAGAGATACCAT 589
Db 517	TGCAAGAGGCTTTTAAAGACACAGCAGCTGAGAACCATAGTGCACGACACACAGC 576
QY 590	GAACTCTATTCAGTGTACCCAGGAAGATGTGGGAACCTTTTCATCACCAGCAAG 649
Db 577	GAGCGCTCTTCAAGTGTACCCAGGAGATGCGGGAAGCACTTCTCTCGCCAGCAG 636
QY 650	CTGAAAACGATGCCAAGGCCACGAGGGCTATGATCAAAAAGATGTTCTTTGG 709
Db 637	CTGAAAACGATGCCAAGGCCACGAGGGCTACTGTGTAAAAGGATGTTCTTTGATG 696
QY 710	GCAAAAACATGAGGAGGAACCTTGTGAACATGTGAGAGAAACCATTAAGAGAAATACTA 769
Db 697	GCAAAAACATGAGGAGGATGCTGTGAACATGTGAGAGAAACCATTAAGAGGACATTAAC 756
QY 770	TGGAAGTATGCGCGGAAACATTTAAACGCAAGATGATCTTAAGCAACATGAAAGCT 829
Db 757	TGGAAGTATGCTCAAGAGATGTTCAAGCGCAGAGATTAACCTTAAGCAACATGAAAGCT 816
QY 830	CATGCCCGGAAAGGATGTATGTGCTGTGCAAGAGAGGCTGTGGAAGAACCTATACT 889
Db 817	CATGCCCGGAAAGGATGTATGTGCTGTGCAAGAGAGGCTGTGGAAGAACCTATACTA 876
QY 890	ACGATGTTAATCTGCAAGGCAATATCTCTCTCTCATGAGAAAGCCCTTTTGG 949
Db 877	ACGATGTTAATCTGCAAGGCAATATCTCTCTCTCATGAGAAAGCCCTTTTGG 936
QY 950	TGTGAACATGCTGCTGTGCGCAAAAACATTTGCAATGAAACAAAGTCTCACTAGGCACTG 1009
Db 937	TGTGAACATGCTGCTGTGCGCAAAATTCGATGAAACAAAGTCTCACTAGGCACTG 996
QY 1010	GTTGTACATGATCTGATCAAGAAATGAAATGAAATGCAATCAAAATCTGCTGAAAAA 1069
Db 997	GTTGTACATGATCTGATCAAGAAATGAAATGAAATGCAATCAAAATCTGCTGAAAAA 1056
QY 1070	CGAGATTTGGCTCTCATCTCAAGATATATCCCTCCAAAAGAAACAGGGCAAGC 1129
Db 1057	CGAGATTTGGCTCTCATCTCAAGATATATCCCTCCAAAAGAAACAGGGCAAGC 1116
QY 1130	TTATCTTTTGTGCA---AAAGGAGAGTCAACCACTGTGTGGAAGCAAGATGCTCTG 1186

Db 1117	TACTCTTCCCTTAACGCCAGCGCAGAGTCCAGCAGCAGCCCAAGGCGCCAGCTGCCCGG 1176
QY 1187	ACAGTTGACGATCTTACCTTGCTTA 1212
Db 1177	CCAGCGGCTTACTACTGCTGCTGCTTA 1202
RESULT 7	
AF391798	1071 bp mRNA linear ROD 31-JAN-2002
LOCUS	AF391798
DEFINITION	Rattus norvegicus transcription factor IIA mRNA, partial cds.
ACCESSION	AF391798
VERSION	AF391798.1 GI:18448379
KEYWORDS	
SOURCE	Rattus norvegicus.
ORGANISM	Rattus norvegicus. Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE	1 (bases 1 to 1071)
AUTHORS	Hanas,J.S., Hocker,J.R., Cheng,Y.G., Lerner,M.R., Brackett,D.J., Lightfoot,S.A., Hanas,R.J., Madhusudan,K.T. and Moreland,R.J.
TITLE	CDNA cloning, DNA binding, and evolution of mammalian transcription factor IIA
JOURNAL	Gene 282 (1-2), 43-52 (2002)
MEDLINE	21673967
PUBMED	11814676
REFERENCE	2 (bases 1 to 1071)
AUTHORS	Hanas,J.S., Hocker,J.R., Lerner,M.R., Brackett,D.J., Lightfoot,S.A., Hanas,R.J., Madhusudan,K. and Moreland,R.J.
TITLE	Submitted (11-JUN-2001) Biochemistry/Molecular Biology, University of Oklahoma Health Science Center, 800 Research Parkway, Suite 446, Oklahoma City, OK 73104, USA
JOURNAL	Location/Qualifiers
FEATURES	1..1071
source	/organism="Rattus norvegicus" /strain="Sprague-Dawley" /db_xref="taxon:10116" /tissue_type="brain"
CDS	<1..1011 /note="FIIIA: Cys2His2 zinc finger protein; similar to Xenopus laevis transcription factor IIA Swiss-Prot Accession Number P03001" /codon_start=1 /product="transcription factor IIA" /protein_id="FAL69685.1" /db_xref="GI:18448380" /translation="AARAGALPSRFCTSPDQSAZYKAMKLDHLKNGENSEFVD YEGGKAFIDYHLSRIHLHTGERPVCADNGCQKSTKSNLKHIERKHENPKO YVNEBCKRAFKRHOOLRTHQOHNEDLFCCTHEGGGKHFASPSRLKRGVHEDY LCQNGSEFVKTVELIKITREAHKEEYVCTQGMFPRKHLKQIMKHTIAPEDVYR CPREGCAFTYTVENLQSHITLSFHEERKPFVCEHAGCGKTPAMQSLMRISVVDPK KRMKLKVRPPRRHSLASRLSGVVPKRGQRPDCLSPSTSSSPKATMLAPALLT VH"
BASE COUNT	294 a 283 c 290 g 204 t
ORIGIN	
Query Match	57.0%: Score 691; DB 10; Length 1071;
Best Local Similarity	80.9%: Pred. No. 6.1e-145;
Matches 805: Conservative 0: Mismatches 190: Indels 0: Gaps 0:	
QY 218	GGCCTTCCAGGAGGTTGATCGCTCTCCCTGACGTGAGCGCCATTAAGCAAGCC 277
Db 16	GGCCTTCCAGGAGGTTGATCGCTCTCTCTCTGACGTGAGCGCTTACCAACAAAGCC 75
QY 278	TGGAAGCTTGACCCCGACCTGTGCAAGCACAGGGGAGAGACCATATTTGTTGTACTAT 337
Db 76	TGGAAGCTTGAGAGCCCGACCTGTGAAGCACAGGGGAGAGGTCGTTGCTGTGTGACTAT 135
QY 338	GAAAGGTGTGGAAGCGCTTCATCAGGAGCTACCATGTGAGCGGCACATTCGACTAC 397
Db 136	GAGGCTGTGGAAGCGCTTCATCAGGAGCTACCATGTGAGCTGACACACATCTGATTCAC 195

OY	398	ACAGGAAACCCTTTTGTTCAGCACTGGCTGTATCAAAAATTCAACAACAAA	457
Db	196	ACCGAAMAAAGCATTITTTGTGTGCAGTAATGGCTTATTCAGAAATTCAGCACAAAG	255
OY	458	TCAAACCTTGAAACAATTTTGACGCCAACATGAAATTCAGAAAAACAATATATATGC	517
OY	518	AGTTTTGAACACTGTAGAAGACCTTTTAAAGACATCGACAGCTGAATAATCCATCATAGTGC	577
Db	316	AATTTTGAACGTTGCAAGAAAGGCTTTTAGAAGACACGACGAGCTTGAGAACCCATCATAGTGC	375
OY	578	CACGATACCAATTAACCTGTATTCGAAGTGTACCCGAGAAAGATATGGGAACCACTTTGCCA	637
Db	376	CAGCACCAATATAGCCACTCTTCAGGTGTACCAAGGAGATATGGGAAGCACTTTGGCC	435
OY	638	TCACCCAGCAAGCTGAAAGACAGATGCCAAAGCCCCACGAGGCTATGTATGTTCAAAAAGCA	697
Db	436	TCCCCCACGAGCTGAAAGCGATGGGAAAGTTTCACGAGGCTCATCTATGTCAAAAAGGCA	495
OY	698	TGTTCCCTTTGACAAAACATGAGACGGAACCTCTGAAACATGTAGAGAAACCCATATA	757
Db	496	TGTTCCCTTTGGGAAAAAACGTGCACAGCTCTTCGAAAGCACAGAGAGAAAGCCCATATA	555
OY	758	GAGCAATATCTATGTACAGTATGCCGAAAAACATTTAAAGCAAAAGATATACCTTAAGCAA	817
Db	556	GAGAGGTGACCTGCACAGATGGCCAGAAAGATGTCAAGCCCMAAGATATCAATAGCACAG	615
OY	818	CACATGAAACACTCATGCCCAAGAAAGGATGTATGTCGCTCTCCAAAGAAAGGCTGTGA	877
Db	616	CACATGAAGACTCAGCTCCGGAGGAGTGTATCCGCTGCCGCGGGGAAAGGCTGTCCA	675
OY	878	AGAACCTATCTACTGTGTGTTTAATCTCCAAAGCCATATCTCTCTCTTCATGAGAAAGC	937
Db	676	AGGACCTACACACCGCTGTTCAACTGCAGAACCCACATTCCTCTTCACAGAGAAAGG	735
OY	938	CGCCCTTTTGTGTGTGAACATCGTGGCTGTGGCAAAACATTTGGCAATGAAACAAAGTCTC	997
Db	736	CGCCATTTTGTGTGATGATGCTGGCTGTGGCAAGACGTTTGGATGAAACAAAGTCTC	795
OY	998	ACTAGGATCTCTTTATCATGATCTCTGCAAGAAAGAAATGAAGTCACAAAGTCACAAAAA	1057
Db	796	ATGAGGCACAGTCTCTGTGCACATCTCTGACAGAAAGAGATGAAGCTTCAGATTTAGACCA	855
OY	1058	TCTCGTAAAAAAGGAGTTGGGCTCTCATCTCACTGATGATVATATCCCTCCCAAAAGGAAA	1117
Db	856	CTCTGGGAGAGACGACAGCTTGGCTCTCGCCCTCACACGGGTACGTCCCTTAAGGGGAAA	915
OY	1118	CAAGGCGAAGCGTATCTTTGTGTCAAAAACGAGAGTCAACCCACTGTGGGAAAGCAAG	1177
Db	916	CAGAGCGCGACACTGCTCTTGGCTTAACGACACAGATGCTCACACAGCCACAGAGGCCACG	975
OY	1178	ATGCTCTGCACAGTTCAGTACTTACCTTGGCTTA	1212
Db	976	ATGCTCTGCCACAGCTCGTTACTTACTGTCACTA	1010
RESULT B	BC032292	996 bp	mRNA linear ROD 07-AUG-2003
LOCUS	BC032292		
DEFINITION	Mus musculus, general transcription factor ttf A, clone MGC:40923		
ACCESSION	BC032292		
VERSION	BC032292.1	GI:21595507	
KEYWORDS	MGC.		
SOURCE	house mouse.		
ORGANISM	Mus musculus.		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
AUTHORS	Mommalia; Eutheria; Rodentia; Sciurognathhi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 996)		
	Strausberg, R.		
	Direct Submission		

REMARK	JOURNAL
<p>Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA</p> <p>NIH-MGC Project URL: http://mgc.ncl.nih.gov</p> <p>Contact: MGC help desk</p> <p>Email: cgapbs-remail.nih.gov</p> <p>Tissue procurement: Gilbert Smith, Ph.D.</p> <p>cDNA Library Preparation: Life Technologies, Inc.</p> <p>cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN)</p> <p>DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),</p> <p>Galthersburg, Maryland;</p> <p>Web site: http://www.nisc.nih.gov/</p> <p>Contact: nisc-mgc@nsl.nih.gov</p> <p>Akheri, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kong, P., Latic, P., Legaspi, R., Maduro, O.L., Mastello, C., Maskell, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantirlop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.</p>	<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILN at: http://image.llnl.gov</p> <p>Series: IRMK plates: 65 Row: 1 Column: 4</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.</p> <p>Location/Qualifiers</p> <p>1..996</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/map="FVB/N-3"</p> <p>/clone="MGC:40923 IMAGE:5374268"</p> <p>/tissue_type="Mammary tumor MMTV-LTR/INT3 model. 5 month old mouse. Taken by biopsy."</p> <p>/clone_llb="NCL_CGAP_Mam2"</p> <p>/lab_host="DH10B"</p> <p>/note="Vector: pCMV-Sport6"</p> <p>365..874</p> <p>/codon_start=1</p> <p>/product="general transcription factor III A"</p> <p>/protein_id="AAH3292.1"</p> <p>/db_xref="GI:21595508"</p> <p>/db_xref="LocusID:66596"</p> <p>/translation="MGKTVTELLKHMREAHKEDITCNVCOMRFKRDYLYKOHMKTNAPEDVYKPCROCCGRTYTYTVEINOSHILSFHEKRPFVCEHAGCCKTFAMKOSLKRHSYVHDPPKRMKRLKTVRAPRERSRLASRLSGYFPKPKRQEPDYSLNPASSESSSSPEADLPPTATLTTC"</p>
<p>BASE COUNT</p> <p>319 a 261 c 244 g 172 t</p> <p>ORIGIN</p>	
<p>Query Match</p> <p>Best Local Similarity 78.9%; Score 560.6; DB 10; Length 996;</p> <p>Matches 681; Conservative 0; Mismatches 179; Indels 3; Gaps 1;</p>	
<p>353 GCCCTCATCGAGGACATCATGTGAGCGCCACCATCTGTGACATCAGACAGAGAAAGCCG 412</p> <p>Db 11 GCCTCATCAGAGACATCAGCATGTGAGCGCGGAGTGTCTGATTCACACCGGGGAAAGCCG 70</p>	
<p>413 TTGTGTTGGCGACCGCATGCGCTGTGATCAAAAATTCAACACAAATTCAACTTGAAGAA 472</p> <p>Db 71 TTGTGTTGGCGACATGATGCGCTGATTCAGAAATTCACAAACAAATTCAACTTGAAGAA 130</p>	
<p>473 CATTTTGAACGCAACATGAATATCAACAAACAAATATATATGACAGTTTGAAGACTGT 532</p> <p>Db 131 CACATTTGAACGCAACATGAAGAACCCACAAACAAACAGTATGTGTGACATTATGAGGCTTC 190</p>	
<p>533 AAGAAGACCTTTAAGAACATCAGACGCTGAAATTCATCATCAGTGCAGCATACCAATGA 592</p> <p>Db 191 AAGAAGACCTTTAAGAAGCACGACGCTGGAACCACTATAGTGCAGCACACACACG 250</p>	

OY 593 CCGTCATCAAGTGTACCCAGGAAGATGTGGAAACACTTGATCACCACCAAGCTG 652
 Db 251 CCACCTCTCAAGGTATCCACAGAGGATGCGGGAAACCACTTTGCTGCCACAGGCTG 310
 OY 653 AAACGACATGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGSCA 712
 Db 311 AAACGGCATGGGAAATTCCTACAGAGGCTACCTGTGTCAAAAAGGATGTTCTTTCAAGGGA 370
 OY 713 AAAACATGACGGAACCTTCGTGAACATGTGAGAGAAACCCATAAGAGSAAATACATATG 772
 Db 371 AAAACGTGGACAGAGCTTCTGAACACATGAGAGAGGCCATAAAGAGSACATTAACCTTCG 430
 OY 773 GAAATATGCCGGAAAAACATTTAAACGCAAAATTTACCTTTAAGCAACACATGAAAACTCAT 832
 Db 431 AAGCTATGTGAGAGGTGTTCGAAGCGCAGAAATTTACCTTTAAGCAACACATGAAGACTGCAC 490
 OY 833 GCCCCGAAAAAGGATGTATGTGCTGTCCAGAGAGAGGCTGTGGAGAAACCTATTAAT 892
 Db 491 GCCCCGAAAAAGGATGTATGACGCTGTCCGCGGCAAGGCTCGGAAGAACCTACACAAACC 550
 OY 893 GTCGTTAAATCTCCAAAGCCATTTCTCTCTCTTCATATGAGAAACGCCCTTTTGTGTGT 952
 Db 551 GTGTCAACCTGACAGGCCATTTCTCTCTCTTCACAGGAGAAAAAGGCCCTATTTGTGTGT 610
 OY 953 GAACATGTGGCTGTGGCAAAACATTTSCAATGAAACAAACCTCATAGACATGCTGTT 1012
 Db 611 GACACAGCTGGCTGTGTGGCAAGACATTCGCAATGAAACAGAGTCTCATAGAGCACACTGTC 670
 OY 1013 GTACATGATCTCTACAGAAGAAATGAAAGTCAAAAGTCAAAAATATGTGTGAAAAACG 1072
 Db 671 GTCCACAGTCCGACAGAAGAGAGATGAAGTCAAAAGTAAAGAGGCCCTGTGGAGAAACGC 730
 OY 1073 AGTTTGCCCTCTCATCTCAGTGGATATATCCCTCCCAAAAGAAACAGGGCAAGGCTTA 1132
 Db 731 AGCTTGGCCCTCTGCTCAGTGGATCTTCCTCTCTTAAGAGAAACAGAGCCGCACTAC 790
 OY 1133 TCTTTTGTGCA--AAACGAGAGTACACCAACATGTGTGAGAGCAAGATGCTGTGACA 1189
 Db 791 TCTGTGGCTTAAGCCACAGGCGACAGTCCAGACAGCAAGTCCAGAGGCCGACGCTGCCCGCCA 850
 OY 1190 GTTTCAGATCTTAACCTTGCTTA 1212
 Db 851 GCCACCTTATCTCACTGTCTGCTTA 873

RESULT 9	LOCUS	AC004739/c	97979 bp	DNA	linear	PRI 03-FEB-2000
DEFINITION		AC004739				
ACCESSION		AC004739				
VERSION		AC004739.1				
KEYWORDS		HTG.				
SOURCE		Homo sapiens.				
ORGANISM		Homo sapiens.				
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						

Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@watsn.wustl.edu
Summary Statistics
Center project name: H_D0531G15

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping project (Eric D. Green, director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GT/CH7>, send eric.green@nhgri.nih.gov, or see <http://www.genome.wustl.edu/gsc>

SOURCE INFORMATION: This clone was derived from human PAC library RCI-4, prepared by *Pieter de Jong* and coworkers at Roswell Park Cancer Institute, using the method described by *Ioannou et al.*, *Nature Genetics* 6:84-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of RP4-531G15;
The actual end is at base position 97979 of RP4-531G15.

This clone contains STS SWSS2758 (MID:g1113567).

1. .97979

```

Repeat_region
/db_xref="taxon:9606"
/chromosome="7"
/map="p21"
/clone="Rp4-51615"
/clone_1lb="RPC1-4"
1..62
/rpc_family="Alu"
137..179
/rpc_family="L2"
660..831
/rpc_family="MIR"
3238..3572
/rpc_family="Alu"
3868..4067
/rpc_family="MIR"
7085..8157
/rpc_family="L1"
8397..8714
/rpc_family="Alu"
8718..9750
/rpc_family="L1"
9751..12094
/rpc_family="L1"
12095..12372
/rpc_family="Alu"
12373..12436
/rpc_family="CA1"

```

```
repeat_region 12437..12842  
/rpt_family="L1"  
repeat_region 12888..13001  
/rpt_family="L1"  
repeat_region 13003..13358  
/rpt_family="MERL"  
repeat_region 13365..14214  
/rpt_family="L1"  
repeat_region 14264..14417  
/rpt_family="L1"  
repeat_region 14415..14729  
/rpt_family="L1"  
repeat_region 14816..15026  
/rpt_family="MERL_type"  
repeat_region 15648..15729  
/rpt_family="CAFA)n"  
repeat_region 16041..16362  
/rpt_family="L1"  
repeat_region 16679..16732  
/rpt_family="L1"  
repeat_region 17025..17315  
/rpt_family="Alu"  
repeat_region 17465..17685  
/rpt_family="L1"  
repeat_region 18044..18633  
/rpt_family="L1"  
repeat_region 18630..18750  
/rpt_family="L1"  
repeat_region 18760..19497  
/rpt_family="L1"  
repeat_region 19508..19692  
/rpt_family="L1"  
repeat_region 19702..19814  
/rpt_family="L1"  
repeat_region 21491..21543  
/rpt_family="L2"  
repeat_region 21987..24424  
/rpt_family="L1"  
repeat_region 24425..24720  
/rpt_family="Alu"  
repeat_region 24722..24821  
/rpt_family="L1"  
repeat_region 25383..25745  
/rpt_family="L1"  
repeat_region 29202..29272  
/rpt_family="MERL_type"  
repeat_region 30338..30386  
/rpt_family="(CA)n"  
repeat_region 30984..31045  
/rpt_family="MERL_type"  
repeat_region 31041..31398  
/rpt_family="MERL_type"  
repeat_region 32178..32319  
/rpt_family="MERL"  
repeat_region 32475..32515  
/rpt_family="L2"  
repeat_region 32589..32646  
/rpt_family="L2"  
repeat_region 32940..33324  
/rpt_family="Alu"  
misc_feature 33843..33929  
/note="match to EST AA928509 (NID:93076800) om17907.s1"  
repeat_region 35891..36381  
/rpt_family="L2"  
repeat_region 37910..38123  
/rpt_family="MERL_type"  
repeat_region 38302..38768  
/rpt_family="L1"  
repeat_region 42711..43026  
/rpt_family="L1"  
repeat_region 43023..43418
```

```
repeat_region /rpt_family="L1"  
43552..43886  
repeat_region /rpt_family="L1"  
43847..44539  
repeat_region /rpt_family="L1"  
44534..45153  
repeat_region /rpt_family="L1"  
45132..45245  
repeat_region /rpt_family="L1"  
45246..45667  
repeat_region /rpt_family="MERL"  
45674..45869  
repeat_region /rpt_family="L1"  
45871..46089  
repeat_region /rpt_family="Alu"  
46091..47111  
repeat_region /rpt_family="L1"  
47186..47354  
repeat_region /rpt_family="MERL_type"  
47529..47577  
repeat_region /rpt_family="MERL_type"  
47673..48037  
repeat_region /rpt_family="L2"  
48067..48494  
repeat_region /rpt_family="L2"  
51739..51788  
repeat_region /rpt_family="(CA)n"  
52362..52665  
repeat_region /rpt_family="L1"
```

Query Match 38.8% Score 470.2; DB 9; Length 97979;

Best Local Similarity 76.8%; Pred. No. 66-95; Matches 655; Conservative 0; Mismatches 158; Indels 40; Gaps 5;

```
QY 329 TGTGACTATGAAGGGTGTGGCAAGGCTTCATCAGGACTACCTGTAGCCGCACANT 388  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79437 TGTGACTATGAAGGGTGTGGCAAGGCTTCATCAGGACTACCTGTAGATCGCCACATC 79378  
QY 389 CGAGCTCACACAGAGAAAGCCGTTTGTTCGACGCCACGTGGTGTATCAA----- 443  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79377 CTGATTCACATTCGAGAAACCTTTATTTCGACGCTGTGTATATAAAAAA 79318  
QY 444 -----AATTCACACAAATCAACTTGAAGAAACATTTGAACCCAAATGAAATC 497  
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79317 AAAAAGAAATCCACAGAAATCAACTGGAAGAAACATTTGATCCAAATGAAATC 79258  
QY 498 AACCAAAATATATATGAGCTTTGAAGCTGTAGAAGACCTTAAGAAACATCAGC 557  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79257 TCCAAAAGCAATATATATGCAATTTTGAAGATGTAGAAGACCTTAAGAAATCAGC 79198  
QY 558 ACCTGAATTCATCAGTCCACATACCAATCACTCTATTCAAGTACCCAGAG 617  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79197 ACCTGAATTCATCAGTCCACATGAAATAAACCCCTATTCAAGTTAACCCTTGAG 79138  
QY 618 GATGTGGGAACACTTCATCACCACGACGTGAAGCAATGCCAAGGCCACAGAG 677  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79137 GATGTGGAAACACTTTTTCACCTCCGACGCTGGAGAGACATGACAGTTCAATCAG 79078  
QY 678 GCTATGTATCAAAAAGCATGCTTGTGGCAAAACATGACGCAACTTGTGAAC 737  
| ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||  
DB 79077 AATATATATGTCMAATGACGATCTTTGTGCGAAGAAATGAC----- 79032  
QY 738 ATGTGAGGAAGCAATCAAGAGAAATACATGATGAGATCCGGAATCAATTAAAC 797  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79031 ---AGAGGAATTCATGAAAGGAATA--ACATGATGTATCCGAAACATTCMAAT 78977  
QY 798 GCAAGATTACCTTAAGCAGACATGAAGCAATGCCCCAGAAAGGATGTATGCGCT 857  
| ||||| | ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 78976 GTGAGATTATCTCAGATTAACCTGTGAACCTGATGCTCCTGAAAGAGTATGTCAAC 78917  
QY 858 GTCCAGAGAGGCTGTGGAAGACATATCTGCTGTATATCCAAAGCCATATCC 917  
||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 78916 ATCCAGAGAGACTGCAAGAG-----ACTGATTTAATCTCCAGAGCTGTCTTC 78866
```


TITLE Direct Submission
JOURNAL Submitted (30-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 111722)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 111722)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 19, 2002 this sequence version replaced g11796493.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0515K14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Fritngen,E., Tatemio,M., Catanese,J.J., and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is CTD-2105K18, 2000 bp overlap. The clone sequenced to the right is RP11-533K11, 2000 bp overlap. Actual start of this clone is at base position 141753 of CTD-2105K18; actual end is at base position 16829 of RP11-533K11.

Polymorphisms have been identified between AC016444 and AC074390. Data from AC016444 was used to finish this clone, AC074390.
Location/Qualifiers
1. 111722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-515K14"

repeat_region	/clone_11b="RPC1-11"
repeat_region	1. 249
repeat_region	/rpf_family="MALR"
repeat_region	250. 640
repeat_region	/rpf_family="MALR"
repeat_region	1650. 2014
repeat_region	/rpf_family="ERV1"
repeat_region	2015. 3233
repeat_region	/rpf_family="L1"
repeat_region	3234. 3578
repeat_region	/rpf_family="ERV1"
repeat_region	3586. 3835
repeat_region	/rpf_family="ERV1"
repeat_region	4560. 4873
repeat_region	/rpf_family="ERV1"
repeat_region	4880. 4975
repeat_region	/rpf_family="GA-rich"
repeat_region	4991. 5133
repeat_region	/rpf_family="ERV1"
repeat_region	5577. 5801
repeat_region	/rpf_family="L1"
repeat_region	6687. 6982
repeat_region	/rpf_family="ALU"
repeat_region	7392. 7449
repeat_region	/rpf_family="AT-rich"
repeat_region	7495. 7564
repeat_region	/rpf_family="AT-rich"
repeat_region	7559. 8728
repeat_region	/rpf_family="L1"
repeat_region	8746. 9038
repeat_region	/rpf_family="ALU"
repeat_region	9017. 9053
repeat_region	/rpf_family="(TAAA)n"
repeat_region	9061. 9362
repeat_region	/rpf_family="ALU"
repeat_region	9372. 10143
repeat_region	/rpf_family="L1"
repeat_region	9934. 9963
repeat_region	/rpf_family="(TG)n"
repeat_region	10144. 10454
repeat_region	/rpf_family="ALU"
repeat_region	10427. 10454
repeat_region	/rpf_family="(A)n"
repeat_region	10455. 11140
repeat_region	/rpf_family="L1"
repeat_region	11487. 11960
repeat_region	/rpf_family="L2"
repeat_region	12112. 12305
repeat_region	/rpf_family="L2"
repeat_region	12306. 12615
repeat_region	/rpf_family="ALU"
repeat_region	12591. 12615
repeat_region	/rpf_family="(A)n"
repeat_region	12616. 12766
repeat_region	/rpf_family="L2"
repeat_region	13024. 13088
repeat_region	/rpf_family="(TA)n"
repeat_region	13613. 13641
repeat_region	/rpf_family="AT-rich"
repeat_region	14281. 14328
repeat_region	/rpf_family="(TA)n"
repeat_region	14329. 15028
repeat_region	/rpf_family="MER2_type"
repeat_region	15050. 15494
repeat_region	/rpf_family="MER2_type"
repeat_region	15445. 15464
repeat_region	/rpf_family="(T)n"
repeat_region	15495. 16698
repeat_region	/rpf_family="L1"
repeat_region	17544. 17927
repeat_region	/rpf_family="L1"
repeat_region	19584. 19618
repeat_region	/rpf_family="MALR"

FEATURES

Source

```

repeat_region      21646..21769
                    /rpl_family="MIR"
repeat_region      21860..22043
                    /rpl_family="MER1_type"
repeat_region      22044..22227
                    /rpl_family="MALR"
repeat_region      22428..22467
                    /rpl_family="MER1_type"
repeat_region      23300..23664
                    /rpl_family="MALR"
repeat_region      23870..23910
                    /rpl_family="(TAA)n"
repeat_region      24405..24635
                    /rpl_family="L1"
repeat_region      24636..24901
                    /rpl_family="MALR"
repeat_region      25481..25504
                    /rpl_family="AT_rich"
repeat_region      26543..26672
                    /rpl_family="AT_rich"
repeat_region      26769..26911
                    /rpl_family="MER1_type"
repeat_region      26928..27295
                    /rpl_family="CR1"
repeat_region      27653..27678
                    /rpl_family="MALR"
repeat_region      27934..28093
                    /rpl_family="AT_rich"
repeat_region      29167..29194
                    /rpl_family="MER1_type"
repeat_region      29191..29309
                    /rpl_family="AT_rich"
repeat_region      29348..29644
                    /rpl_family="Alu"

```

Query Match 37.3% Score 452.6; DB 9; Length 111722;

Best Local Similarity 80.3% Pred. No. 5,6e-91; Matches 685; Conservative 0; Mismatches 139; Indels 29; Gaps 12;

```

209 CCGCCGCCGCCGCTCCGAGAGGTGTCCTGCTCTCCCTGACTGACAGCCCAATTAC 268
    || || || || || || || || || || || || || || || || || || || ||
Db 87601 CCCCTTCACACACTCTTAGAGAGGTGTCCTGCTCTCCCTGACTGACAGCCCAATTAC 87543

269 AGCAAGCTGGAACCTTGACGCCGACCTGTGCAGCAGCAGCGGGGAGAGACATTGTT 328
    || || || || || || || || || || || || || || || || || || || ||
Db 87542 AAAAGGCGCTGGAAGCTTAAAGTGCACCTGT--ATGACATGGGGAAGAACCATTTGTT 87485

329 TGTGACTATGAAGGCTGTGCGAAGGCGCTTCATC-AGGACTACCATCTGAGCGCCACAT 387
    || || || || || || || || || || || || || || || || || || || ||
Db 87484 TATGACTGGAAGGCTGTGCGAAGGCGACCTCAAGAGACTAATCTGAGTTGTCATGT 87425

388 TGTGACTGCAGAGAGAAAGCCGTTTGTGTGCGACGACCTGCGCTGATC-AAAAT 446
    || || || || || || || || || || || || || || || || || || || ||
Db 87424 CCGTA-TCATCTGCGAGAGCCATATATGTTGTGCGAGCTAGTGTGATCTAAATAAT 87366

447 TCACACAAATCAAACTTGAAGAAACATTTTGAACGCAAAACATGAAAAATCAACAAAA 506
    || || || || || || || || || || || || || || || || || || || ||
Db 87365 TCACACAAATCAAACTTGAAGAAACATTTTGAACGCAAAACATGAAAAATCAACAAAA 87306

507 A-----ATATATATGCACTTTTGAAGACTGTAAAGAGCTTAAAGAAATCAGCAGCT 561
    || || || || || || || || || || || || || || || || || || || ||
Db 87305 AAAATATATCAATATGCAATTTTGAAGGTTGGAGAGAGACCTTACG-ACATCAGCAGCT 87247

562 GAAATTCATCAGTGCAGCATACCAATGAACCTCTATTCAGTGTACCCAGGAGAGATG 621
    || || || || || || || || || || || || || || || || || || || ||
Db 87246 GAAATTCATCAGTGCAGCATACCAATGAACCTCTATTCAGTGTACCCAGGAGAGATG 87187

622 TGGGAAAGCCTTGGATACCCGAGAGCTGAAGGACATGCGAGCCGACAGAGGCTA 681
    || || || || || || || || || || || || || || || || || || || ||
Db 87186 TGGGAAAGCCTTGGATACCCGAGAGCTGAAGGACATGCGAGCCGACAGAGGCTA 87134

682 TGTATGTCAAAAAAGATGTCCTTTGTGGCAAAAAATGAGCGAATTCGAAACATGT 741
    || || || || || || || || || || || || || || || || || || || ||
Db 87133 TATATGTCAAAAAAGATGTCCTTTGTGGTGAAGAGATGAAGAGTTTGAACATGT 87074

```

```

Oy 742 GAGAGAAACCCATAAGAGGAATACTATGTGAAGTATGCCGAAACATTTAAACGCA 801
    || || || || || || || || || || || || || || || || || || || ||
Db 87073 GAGAGAAACCTTATTAAGAAACACATATGAGCAGCGCAAAAAACATTTAAAC---- 87018

802 ACATTACCTTTAAGACACATGAAACCTCATAGCCCGAGAAAGGATGTATGCGCTGCC 861
    || || || || || || || || || || || || || || || || || || || ||
Db 87017 -ATTATCTTAAGACACATATGAACTCATAGCCCGAGAAAGGATAT-TTTGACATAC 86961

862 AAGAGAAAGCTGTGGAAGAACCTTACTACTGTGTTAATCTGCAAGCATATCTCTC 921
    || || || || || || || || || || || || || || || || || || || ||
Db 86960 AAGAGAAAGCTGTGGAAGAACCTTACTACTGTGTTAATCTGCAAGCATATCTCTC 86901

922 CTTCATGAGAAACCCGCTTTGTGTGGAACATGCTGCTGCGCAAAACATTTGC 981
    || || || || || || || || || || || || || || || || || || || ||
Db 86900 CAT--GAGAAAAAAGCTGTCATTTTGCATGTGAACATGTGGCTGGCAAAACATTTGC 86843

982 AATGAACAAAGCTCTACATAGCATGCTGTTTACATGATCCCGACAAAGAAATGAA 1041
    || || || || || || || || || || || || || || || || || || || ||
Db 86842 AATGAACAAAGCTCTTCTTACGATGCTGTTTACAT-ATTGACAAAGAAATGAA 86784

Oy 1042 GCTCAAGTCANA 1054
    || || || || ||
Db 86783 GCTCAGAGTAATA 86771

```

RESULT 12
AC016444
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-384114 map 11, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
AC016444
VERSION
AC016444.3 GI:10045722
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
1 (bases 1 to 206622)
Birren,B., Linton,L., Nussbaum,C. and Lander,E.
Homo sapiens chromosome 11, clone RP11-384114
JOURNAL
REFERENCE
2 (bases 1 to 206622)
Unpublished

Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeRellano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardina,S., Grant,G., Hago,B., Heath,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGuck,A., McKernan,K., McLaughlin,D., Meldrum,J.,
Morrow,J., Naylor,D., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirelli,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (26-Nov-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6778507.
All repeats were identified using RepeatMasker:

Smilc,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MTBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L5006
Center clone name: 384_I_14
Summary Statistics

```

Sequencing vector: M13: M77815, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 190829 bases at least Q40
Consensus quality: 198597 bases at least Q30
Consensus quality: 202013 bases at least Q20
Insert size: 213000; agarose-fp
Insert size: 205122; sum-of-contigs
Quality coverage: 4.2 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 13492: contig of 13492 bp in length
* 13493 13592: gap of 100 bp
* 13593 14758: contig of 1166 bp in length
* 14759 14858: gap of 100 bp
* 14859 16187: contig of 1329 bp in length
* 16188 16287: gap of 100 bp
* 16288 18628: contig of 2341 bp in length
* 18629 18728: gap of 100 bp
* 18729 22545: contig of 3817 bp in length
* 22546 22645: gap of 100 bp
* 22646 27333: contig of 4688 bp in length
* 27334 27433: gap of 100 bp
* 27434 32718: contig of 5285 bp in length
* 32719 32818: gap of 100 bp
* 32819 37975: contig of 5157 bp in length
* 37976 38075: gap of 100 bp
* 38076 48059: contig of 9984 bp in length
* 48060 48159: gap of 100 bp
* 48160 58135: contig of 9976 bp in length
* 58136 58235: gap of 100 bp
* 58236 69592: contig of 11357 bp in length
* 69593 69692: gap of 100 bp
* 69693 80834: contig of 11142 bp in length
* 80835 80934: gap of 100 bp
* 80935 102411: contig of 21477 bp in length
* 102412 102511: gap of 100 bp
* 102512 127924: contig of 25413 bp in length
* 127925 128024: gap of 100 bp
* 128025 162338: contig of 34314 bp in length
* 162339 162438: gap of 100 bp
* 162439 206622: contig of 44184 bp in length.
*
* Location/Qualifiers
*
* 1. 206622
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="11"
* /map="11"
* /clone="RP11-384114"
* /clone_1fb="RPC1-11 Human Male BAC"
*
* 1. 13492
* /note="assembly-fragment
* clone_end:sp5
* vector_side:left"
*
* misc_feature
* 13593..14758
* /note="assembly-fragment"
*
* misc_feature
* 14859..16187
* /note="assembly-fragment"
*
* misc_feature
* 16288..18628
* /note="assembly-fragment"
*
* misc_feature
* 18729..22545
* /note="assembly-fragment"
*
* misc_feature
* 22646..27333
* /note="assembly-fragment"
*
* misc_feature
* 27434..32718

```

	misc_feature	/note="assembly_fragment"	32819..37975	
	misc_feature	/note="assembly_fragment"	38076..48059	
	misc_feature	/note="assembly_fragment"	48160..58135	
	misc_feature	/note="assembly_fragment"	58236..69592	
	misc_feature	/note="assembly_fragment"	69693..80834	
	misc_feature	/note="assembly_fragment"	80935..102411	
	misc_feature	/note="assembly_fragment"	102512..127924	
	misc_feature	/note="assembly_fragment"	128025..162338	
	misc_feature	/note="assembly_fragment"	162439..206622	
	misc_feature	/note="assembly_fragment"	clone_end:17	
	misc_feature	vector_side:right"		
BASE COUNT	68410 a	34710 c	35722 g	66279 t 1501 others
ORIGIN				
Query Match	37.3%	Score 452.6:	DB 2:	Length 206622:
Best Local Similarity	80.3%	Pred. No. 6.2e-91:		
Matches 685:	Conservative	0: Mismatches 139:	Indels 29:	Gaps 12:
OY	209	CCGCCGCCCGCGCTCCCAAGAGGTTGATCTGCTCTCTGCTGAGCTGCAGCCCAATTAC	268	
Db	36308	CCCCCTCACACACTCTTGGAGAGGTCATCTCTCTCC-CAGGTGCGACCACCTTAC	36366	
OY	269	AGCAAAAGCTGGAAGCTTGACGCGCCGCTGTCAGACAGCGGGGAGAGCCATTGGTT	328	
Db	36367	AAAAGGCTGGAAGCTTATGTGACACCTGT--AATCAGATGGGGAGAAACCATTTGGTT	36424	
OY	329	TGTACTATGAGAGGTGTGGCAAGGCTTTCATC-AGGACTACCATCTGAGCGCCACAT	387	
Db	36425	TATGACTGTGAAGGGTGTGGCAAGACCCACCTCAAGAGACATACATCTGAGTTGTCTGT	36484	
OY	388	TCTACTCAGACAGAGAGAAAGCCGTGTTGTTGTGCAGCCACTGGCTGTGATC-AAAAAT	446	
Db	36485	CCTGA-TCATATGTCAGACAGCAAGCATATGTTGTGCAGCTAGTAGCTGTGATCAAAAT	36543	
OY	447	TCACACCAAAANTCAACTTGAAGAAACATTTTGAACGCCAAACATGAAATCAACAAAC	506	
Db	36544	TCACACCAAAATCAACTTGAAGAGCATTTTACAGGCAAAACATGAAACCGCAAAAAA	36603	
OY	507	A-----ATATATATGCGATTTTGAAGACTGTAAGAGACCTTTAAGAAACATCGACGT	561	
Db	36604	AAAATATATCATATATGCAATTTTGAAGTGTGAAGAAAGACCTTTACG-AACATCAGCAGCT	36662	
OY	562	GAAATTCATCAGTCGCGCAGCATTCACATGACACCTCTATTCAAGTGTACCCAGAGAGATG	621	
Db	36663	GAAATTCATCAGTCGCGCGATATCCATATGACACCTATCAACTGTACCCAGAGAAATCG	36722	
OY	622	TGGGAACACTTGTGCATCACCCAGCAAGCTGAAGACATCCCAAGGCCCGCAGCGAGCTA	681	
Db	36723	TGGGAACATTTTCCCATTCATCCATCCAGC-----AAGCGGCAAGGGCCATGGGGCTC	36775	
OY	682	TCTATGTCAAAAAGAGTGTCTCTTGTGTGGCAAAACATGAGCGGAGACCTCTGAAACATGT	741	
Db	36776	TATATGTCAAAAAGCGGTCTTGTGTGTGTGAACCATGACACAGCTTTGAAACATGTG	36835	
OY	742	GAGAGAAACCCATTAAGAGAGAAATACTATGTGAAGTATGCGCGGAAACATTTAAACGCA	801	
Db	36836	GAGAGAAACCTTATTAAGAAAGAAACAATATGACAGCGCCAGAAACATTTAAAC----	36891	
OY	802	AGATTTCTTTTAAACACACATGAAACATCATGATCCCGCAGAAAGGATGTATGTGCTGTCC	861	
Db	36892	--ATTATCTTAAAGACATGTAGACACTCATATCCCGCAGAAAGGATATAT-TTTGTACATAC	36948	
OY	862	AAAGAAAGGCTGTGGAAGAACCTATATCTGTGTTTATATCTCCAAAGCCATATCTCTTC	921	

OY 1075 TTTGG 1079
Db 124808 TTTGG 124812

RESULT 14

LOCUS AX396111 439 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 326 from Patent W00212328.
ACCESSION AX396111
VERSION AX396111.1 GI:21066858
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0212328-A 326 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1. 439
BASE COUNT 106 a 94 c 99 g 140 t
ORIGIN

Query Match 34.5%; Score 418.4; DB 6; Length 439;
Best Local Similarity 99.8%; Pred. No. 1.1e-83;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 794 AAGCGAAGATTACCTTAAGCAGACATGAAACATGCCCCAGAAAGGATGTATGT 853
Db 439 AAGCGAAGATTACCTTAAGCAGACATGAAACATGCCCCAGAAAGGATGTATGT 380
OY 854 CGCTGTCCAGAGAGGCTGTGGAAGACCTATCTGTGTTAATCTCCAAAGCCAT 913
Db 379 CCCTGTCCAGAGAGGCTGTGGAAGACCTATCTGTGTTAATCTCCAAAGCCAT 320
OY 914 ATCTCTCTCTCCATGAGAGAGCGCCCTTTTGTGTGTAACATGCTGTGGTGCATA 973
Db 319 ATCTCTCTCTCCATGAGAGAGCGCCCTTTTGTGTGTAACATGCTGTGGTGCATA 260
OY 974 ACATTGGCAATGAACAAGTCTCAGTACGATGCTGTGTACATGATCTGACAAGAG 1033
Db 259 ACATTGGCAATGAACAAGTCTCAGTACGATGCTGTGTACATGATCTGACAAGAG 200
OY 1034 AAAATGAGCTCAAAAGTCAAAAATCTCTGAAAAAGAGTTGGCTCTCATCTCAGT 1093
Db 199 AAAATGAGCTCAAAAGTCAAAAATCTCTGAAAAAGAGTTGGCTCTCATCTCAGT 140
OY 1094 GGATATATCCCTCCCAAAAGGAGCAAGGCTTATCTTTGTCAAAACGAGAG 1153
Db 139 GGATATATCCCTCCCAAAAGGAGCAAGGCTTATCTTTGTCAAAACGAGAG 80
OY 1154 TCACCAACTGTGTGGAAGACAAAGATGCTCTGCAGAGTTGCACTACTTACCCTTGCTTA 1213
Db 79 TCACCAACTGTGTGGAAGACAAAGATGCTCTGCAGAGTTGCACTACTTACCCTTGCTTA 20

RESULT 15

LOCUS AX396474 439 bp DNA linear PAT 18-MAY-2002
DEFINITION Sequence 689 from Patent W00212328.
ACCESSION AX396474
VERSION AX396474.1 GI:21067221
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
AUTHORS King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon cancer
JOURNAL Patent: WO 0212328-A 689 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1. 439
BASE COUNT 141 a 99 c 93 g 106 t
ORIGIN

Query Match 34.4%; Score 416.8; DB 6; Length 439;
Best Local Similarity 99.5%; Pred. No. 2.6e-83;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 794 AAGCGAAGATTACCTTAAGCAGACATGAAACATGCCCCAGAAAGGATGTATGT 853
Db 1 AAGCGAAGATTACCTTAAGCAGACATGAAACATGCCCCAGAAAGGATGTATGT 60
OY 854 CGCTGTCCAGAGAGGCTGTGGAAGACCTATCTGTGTTAATCTCCAAAGCCAT 913
Db 61 CCCTGTCCAGAGAGGCTGTGGAAGACCTATCTGTGTTAATCTCCAAAGCCAT 120
OY 914 ATCTCTCTCTCCATGAGAGAGCGCCCTTTTGTGTGTAACATGCTGTGGTGCATA 973
Db 121 ATCTCTCTCTCCATGAGAGAGCGCCCTTTTGTGTGTAACATGCTGTGGTGCATA 180
OY 974 ACATTGGCAATGAACAAGTCTCAGTACGATGCTGTGTACATGATCTGACAAGAG 1033
Db 181 ACATTGGCAATGAACAAGTCTCAGTACGATGCTGTGTACATGATCTGACAAGAG 240
OY 1034 AAAATGAGCTCAAAAGTCAAAAATCTCTGAAAAAGAGTTGGCTCTCATCTCAGT 1093
Db 241 AAAATGAGCTCAAAAGTCAAAAATCTCTGAAAAAGAGTTGGCTCTCATCTCAGT 300
OY 1094 GGATATATCCCTCCCAAAAGGAGCAAGGCTTATCTTTGTCAAAACGAGAG 1153
Db 301 GGATATATCCCTCCCAAAAGGAGCAAGGCTTATCTTTGTCAAAACGAGAG 360
OY 1154 TCACCAACTGTGTGGAAGACAAAGATGCTCTGCAGAGTTGCACTACTTACCCTTGCTTA 1213
Db 361 TCACCAACTGTGTGGAAGACAAAGATGCTCTGCAGAGTTGCACTACTTACCCTTGCTTA 420

Search completed: February 10, 2003, 17:02:32
Job time : 3923.82 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:56:36 : Search time 1896.59 Seconds
(without alignments)
10358.112 Million cell updates/sec

Title: US-09-831-426C-4

Sequence: 1 gtgcgcgcgcgcgcgaag.....cagtactacctgctctaa 1213

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlinu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlinu:*
16: em_estlom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtr:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_procl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	921.4	76.0	1078	13	BM553401 AGENCOURT
2	919	75.8	958	9	AL518337
3	904	74.5	916	9	AL550467
4	877.8	72.4	881	9	AL555777
5	870.6	71.8	1059	13	BM562682
6	862.8	71.1	937	9	AL574318

Result No.	Score	Query Match	Length	DB ID	Description
7	860.4	70.9	930	14	BO650498
8	849.4	70.0	957	9	AL527493
9	844.6	69.6	906	9	AL574698
10	838.8	69.2	893	9	AU117724
11	836	68.9	902	14	BO644275
12	830.2	68.4	933	9	AL578193
13	829	68.3	1089	14	BO278693
14	827.4	68.2	927	12	BO683080
15	827.2	68.2	508	14	BO648620
16	824	67.9	896	12	BO683119
17	822.6	67.8	1110	14	BM809285
18	818.6	67.5	1111	13	BM541851
19	817.4	67.4	892	9	AL518336
20	812	66.9	871	12	BO750799
21	810.4	66.8	890	12	BO697646
22	806.8	66.5	899	14	BO932064
23	801.4	66.1	971	14	BO279125
24	800.4	66.0	842	14	BO430441
25	799	65.9	887	14	BO225045
26	797.2	65.7	880	13	BM047210
27	796.2	65.6	912	12	BE797173
28	796	65.6	910	12	BO281954
29	792.6	65.3	974	13	BI086290
30	791.2	65.2	1027	10	BE562240
31	787.6	64.9	932	9	AL527492
32	784	64.6	883	14	BO436552
33	783.6	64.6	923	13	BM435967
34	781.6	64.4	939	14	BO929358
35	778.8	64.2	879	13	BM041009
36	778	64.1	953	13	BM041678
37	774.8	63.9	801	12	BS519635
38	768.4	63.3	1047	13	BM555545
39	763	62.9	1043	12	BC389788
40	750.4	61.9	874	10	BE410109
41	749.8	61.8	755	13	BI667697
42	738.6	60.9	878	12	BO822097
43	737.8	60.8	753	13	BI759103
44	736.2	60.7	839	13	BI598078
45	732.8	60.4	736	13	BI764623

ALIGNMENTS

RESULT 1
BM553401
LOCUS
DEFINITION
AGENCOURT.6572564 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467230
5', mRNA sequence.
ACCESSION
BM553401
VERSION
BM553401.1 GI:18792097
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 1078)
AUTHORS
NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
CONTACT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LITCM1968 row: 0 column: 07
High quality sequence stop: 673.
Location/Qualifiers
1..1078

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5467230"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/note="Organ: skin; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."
BASE COUNT      301 a      276 c      275 g      223 t      3 others
ORIGIN

```

```

Query Match      76.0% Score 921.4; DB 13; Length 1078;
Best Local Similarity 95.1% Pred. No. 5,9e+194;
Matches 1025; Conservative 0; Mismatches 44; Indels 9; Gaps 7;

```

```

6 GGCACGAGCGGAGGTTGACGAGGAGCGCGGCGCGCGGTTCCGCGACGTG 65
1 GCGCTGCGCGGAGGTTGACGAGGAGCGCGGCGCGCGGTTCCGCGACGTG 60
66 TCTGCGACGTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125
61 TCTGCGACGTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
126 CGCGCGGTCGCGCGACGTCGCGTCGTCCTTGACCATCGCGCGCGTCATG 185
121 CGCGCGGTCGCGCGACGTCGCGTCGTCCTTGACCATCGCGCGCGTCATG 180
186 GCGAGAGTCAGCTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 245
181 GCGAGAGTCAGCTCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
246 TCCCTGACTGCGCGCGCAATTACAGCAAGCGCTTGACGCGCGCGCGCGCG 305
241 TCCCTGACTGCGCGCGCAATTACAGCAAGCGCTTGACGCGCGCGCGCGCG 300
306 ACACGCGGAGAGACCATTTGTTGTGATGATGAAGGTTGCGAAGCGCTTATC 365
301 ACACGCGGAGAGACCATTTGTTGTGATGATGAAGGTTGCGAAGCGCTTATC 360
366 ACTACCATCTGAGCGCGCACATCTGACTCACACAGAGAGAAAGCGTTTGTG 425
361 ACTACCATCTGAGCGCGCACATCTGACTCACACAGAGAGAAAGCGTTTGTG 420
426 CCACTGCGCTGTATCAAAATTCACAAACCAATTCGAAGAAATTTGAAGCGA 485
421 CCAATGCGCTGTATCAAAATTCACAAACCAATTCGAAGAAATTTGAAGCGA 480
486 AACATGAAATTCACAAACCAATTCATGCACTTTTGAAGCTGTAAGAACCTTTA 545
481 AACATGAAATTCACAAACCAATTCATGCACTTTTGAAGCTGTAAGAACCTTTA 540
546 ACAACATCAGAGCTGAAATTCATGAGCGACATACCATGAAGCTTATTCAGT 605
541 ACAACATCAGAGCTGAAATTCATGAGCGACATACCATGAAGCTTATTCAGT 600
606 GTACCCAGGAGAGATGTGGGAACATTTGCATACCCAGCAAGCTGAAGACATGCCA 665
601 GTACCCAGGAGAGATGTGGGAACATTTGCATACCCAGCAAGCTGAAGACATGCCA 660
666 AGGCCAGGAGAGGATATGATGCAAAAGAGATGTCCTTGGGAAACAAACATGAGCG 725
661 AGGCCAGGAGAGGATATGATGCAAAAGAGATGTCCTTGGGAAACAAACATGAGCG 720
726 AACCTCTGAACATGTGAGAGAAACCAATAAAGAGAAATATATGTGATGCGGCA 785
721 AACCTCTG-AACATGTGAGAGAAACCAATAAAGAGAAATATATGTGATGCGGCA 779

```

```

OY 786 AACATTTAAACGCAAAATTAACCTTAAGCAGACATGAAATAATCGCCCGAAGAGG 845
DB 780 NAACATTTAAACGCAAAATTAACCTTAAGCAGACATGAAATAATCGCCCGAAGAGG 838
OY 846 ATGATATGCGCTGTCACAGAGAGAGCTGTGGAGAACCTTACTACTGTTTAATCTCC 905
DB 839 ATGATATGCGCTGTCACAGAGAGAGCTGTGGAGAACCTTACTACTGTTTAATCTCC 898
OY 906 AAGCATATCTCTCTCTTCCATGAGGAAA-GCCGCCCTTTTGTGTGACATGCT--G 962
DB 899 -AAGCATATCTCTCTCTTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957
OY 963 GCTGTGCGCAAAATTTCAATGAAACAAAGTCTCACTAGCATGCTGTGTGA--CATGA 1020
DB 958 CGTGTGCGCAAAATTTCAATGAAACAAAGTCTCACTAGCATGCTGTGTGAATGAT 1017
OY 1021 TCCCTGCAAGAGAGAAAT-GAAGCTCAAGTCGCAAAATTCCTGTAAGAAACGAGATT 1077
DB 1018 CCTGTGCAAGAGAGAAATGAACTTCCAAATTTCCGAAATAACGCGGATT 1075

```

```

RESULT 2
AL518337 958 bp mRNA linear EST 13-FEB-2001
LOCUS AL518337
DEFINITION L1.L.NFI011.NBC1 Homo sapiens cDNA clone CS0DA009Y018 5
ACCESSION AL518337
VERSION AL518337.1 GI:12781830
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 958)
L1.M.B., Gruber, C., Jessee, J. and Polyes, D.
Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

```

```

FEATURES
source
1..958
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DA009Y018"
/clone_lib="L1.L.NFI011.NBC1"
/sex="male"
/tissue_type="neuroblastoma cells"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifestech.com URL : http://fulllength.invitrogen.com"

```

```

BASE COUNT      270 a      244 c      232 g      195 t      17 others
ORIGIN

```

```

Query Match      75.8% Score 919; DB 9; Length 958;
Best Local Similarity 97.5% Pred. No. 2e+193;
Matches 935; Conservative 16; Mismatches 6; Indels 2; Gaps 2;

```

```

OY 16 GAAGCTTACGACGAGGCGGTGGCGCGCGCGCGGTTCCGCGACGTCTCGGACG 75
DB 1 GAAGCTTACGACGAGGCGGTGGCGCGCGCGCGGTTCCGCGACGTCTCGGACG 60
OY 76 TGGACGCGCGCGCTGGCGCTTGAGAGGCGCGCGCGCGCTGTGATCCGCGCGCTGCT 135

```



```

|||||
Db 61 TGGCAGCGCCCTGGCCCTTGGAGGCGCGGCCCTTGATCCGGCCGCTGGT 120
|||||
Qy 136 CGCCAGTCGGTGTGCTCTTGACATCGCCGCTTCATTGCAAGCCGGCGAGATC 195
|||||
Db 121 CGCGAGTGGTGTGTCTCTTGACATCGCCGCTTCATTGCAAGCCGGCGAGATC 180
|||||
Qy 196 AGCTCGAGCCCGCGCGCGCGCTTCCAGAGGTTGATCGCTTCCCTCTACTG 255
|||||
Db 181 AGCTCGAGCMCGCGCGCGCGCTTCCAGAGGTTGATCGCTTCCCTCTACTG 240
|||||
Qy 256 CAGCGCAATATACAGCAAGCCGTGAGAGCTTGACGCGCCTGTGCAAGCAGCGGGA 315
|||||
Db 241 CAGCGCAATATACAGCAAGCCGTGAGAGCTTGAGAGCTTGCAAGCAGCGGGA 300
|||||
Qy 316 GAGACCATTTGTTTGATGATGAGAGGTTGAGAGGCTTCATCGAGGATACCATCT 375
|||||
Db 301 RARACCATTTGTTTGATGATGAGAGGTTGAGAGGCTTCATCGAGGATACCATCT 360
|||||
Qy 376 GAGCGCCCATTTGATGATGATGAGAGGTTGAGAGGCTTCATCGAGGATACCATCT 435
|||||
Db 361 RAGCGCCCATTTGATGATGATGAGAGGTTGAGAGGCTTCATCGAGGATACCATCT 420
|||||
Qy 436 TGTATCAAAATTCACACAAATCAAACTGAAGAACTTTGAAGCAAGCATGAAAA 495
|||||
Db 421 TGTATCAAAATTCACACAAATCAAACTGAAGAACTTTGAAGCAAGCATGAAAA 480
|||||
Qy 496 TCACACAAATTCATATATATGAGTTTGAAGACTGTAAGAGACTTTAAGAAATCA 555
|||||
Db 481 TCACACAAATTCATATATATGAGTTTGAAGACTGTAAGAACTTTAATTAACATCA 540
|||||
Qy 556 GCAGCTGAATATCCATGATGATGATGATGATGATGATGATGATGATGATGATG 615
|||||
Db 541 GCAGCTGAATATCCATGATGATGATGATGATGATGATGATGATGATGATGATG 600
|||||
Qy 616 AGGATGTGGGAAACCTTTGATGATGATGATGATGATGATGATGATGATGATG 675
|||||
Db 601 AGGATGTGGGAAACCTTTGATGATGATGATGATGATGATGATGATGATGATG 660
|||||
Qy 676 GAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 735
|||||
Db 661 GAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
|||||
Qy 736 ACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAGATGATGATGATGATG 795
|||||
Db 721 ACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAGATGATGATGATGATG 780
|||||
Qy 796 ACAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 854
|||||
Db 781 ACAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
|||||
Qy 855 GCTGTCAAGAGAGAGCTGTGAGAGAACTTACTACTGTGTATCTTCAAGCCATA 914
|||||
Db 841 GCTGTCAAGAGAGAGCTGTGAGAGAACTTACTACTGTGTATCTTCAAGCCATA 900
|||||
Qy 915 TCTCTCTTCCATGAGAGAAACCCCTTTTGTGTGAGAACTGTGCTGTGAGAAA 973
|||||
Db 901 TCTCTCTTCCATGAGAGAAACCCCTTTTGTGTGAGAACTGTGCTGTGAGAAA 958
|||||

```

```

AUTHORS  L.J.M.B., Gruber C., Jesse J. and Polayes D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 Evry cedex - France
          Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source    1..916
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone="CS01057YN01"
          /clone_lib="LTI_NFL006_PL2"
          /tissue_type="Placenta"
          /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
          was primed with a NotI-oligo(dT) primer. Five prime end
          enriched, double-stranded cDNA was digested with Not I and
          cloned into the Not I and Eco RV sites of the pCMVSPORT 6
          vector. Library was normalized. Library was constructed by
          a division of Invitrogen 9800 Medical Center Drive
          Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
          Email : filang@lifetech.com URL :
          http://fulllength.invitrogen.com"
BASE COUNT 261 a 238 c 225 g 190 t 2 others
ORIGIN
Query Match 74.5% Score 904; DB 9; Length 916;
Best Local Similarity 99.6%; Pred. No. 4.2e-190;
Matches 904; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 48 GCGGGTTCGCGGAGCTGTCTGTCGACGTCGACGCGCGCGCGCGCGCGCGCGCGCG 107
Db 1 GCGGGTTCGCGGAGCTGTCTGTCGACGTCGACGCGCGCGCGCGCGCGCGCGCGCG 60
Qy 108 CCGGCGCCCTGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 167
Db 61 CCGGCGCCCTGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Qy 168 ACCGCTTCATGAGCGCGCGCGAGACTGACGCTCCGACCGCGCGCGCGCGCGCGCG 227
Db 121 ACCGCTTCATGAGCGCGCGCGAGACTGACGCTCCGACCGCGCGCGCGCGCGCGCG 180
Qy 228 GAGAGTTCATCTGCTCCTCTCCCTGACTGACGCGCAATTACAGCAAGCCGTGAA 287
Db 181 GAGAGTTCATCTGCTCCTCTCCCTGACTGACGCGCAATTACAGCAAGCCGTGAA 240
Qy 288 ACCGCGACCTGTGCAAGCAGACGCGGAGAGACCTTTGTTGTGACTATGAGGGGTG 347
Db 241 ACCGCGACCTGTGCAAGCAGACGCGGAGAGACCTTTGTTGTGACTATGAGGGGTG 300
Qy 348 GCAAGGCTTCATGAGGAGCTACCATCTGAGCGCGCACATTTGCTGACTACAGAGAAA 407
Db 301 GCAAGGCTTCATGAGGAGCTACCATCTGAGCGCGCACATTTGCTGACTACAGAGAAA 360
Qy 408 ACCGCTTTGTTGTGAGGACACTGCTGCTGATCAAAATTCACACAAATCAAACTGA 467
Db 361 AGCGCTTTGTTGTGAGGACACTGCTGCTGATCAAAATTCACACAAATCAAACTGA 420
Qy 468 AGAAACATTTTGAAGCAAGCAATGAAATCAACAAATCAATATATATGCAATTTGAG 527
Db 421 AGAAACATTTTGAAGCAAGCAATGAAATCAACAAATCAATATATATGCAATTTGAG 480
Qy 528 ACAGTAAGAGACCTTTAAGAAATCAAGCAGCTGAAATTCATCACTGCGACATACCA 587
Db 481 ACTGTAAGAGACCTTTAAGAAATCAAGCAGCTGAAATTCATCACTGCGACATACCA 540
Qy 588 ATGAACTCTATTCAAGTGTACCCAGAGAGATGTGGAAACACTTTGATCACCCAGCA 647
Db 541 ATGAACTCTATTCAAGTGTACCCAGAGAGATGTGGAAACACTTTGATCACCCAGCA 600
Qy 648 AGCTGAAGCAGATCGCAAGCGCCACGAGGCGTATGTGCAAAAGATGTTCCCTTG 707

```

```

RESULT 3
AL550467
LOCUS    AL550467 LTI_NFL006_PL2 Homo sapiens cDNA clone CS01057YN01 5
DEFINITION
prime, mRNA sequence.
ACCESSION
AL550467
VERSION  AL550467.1 GI:12887469
KEYWORDS
EST.
SOURCE   human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 916)

```

Db 601 AGCTGAAGCATGCCAAGGCCAGAGGGCTATGTATGTCAAAAAGATGTTCCTTTG 660

QY 708 TGGCAAAAACATGACGACGAACCTTTGTAACATGTGACAGAAACCCATTAAGAAGAAATAC 767

Db 661 TGGCAAAAACATGACGACGAACCTTTGTAACATGTGACAGAAACCCATTAAGAAGAAATAC 720

QY 768 TATGTAAGTATGCGGGAAGAAATTTAAAGCAAAAGTATCTTAAGCAACATGATAA 827

Db 721 TATGTAAGTATGCGGGAAGAAATTTAAAGCAAAAGTATCTTAAGCAACATGATAA 780

QY 828 CTCATATGCCCAAGAAAGGATATGTGCTGTCCAGACAGAGCTGTGGAAGACCTATA 887

Db 781 CTCATATGCCCAAGAAAGGATATGTGCTGTCCAGACAGAGCTGTGGAAGACCTATA 840

QY 888 CTCATATGCCCAAGAAAGGATATGTGCTGTCCAGACAGAGCTGTGGAAGACCTTTTG 947

Db 841 CAATGTGTTTAATCTCAAAAGCATATCTCTCTCTCCATGAGAAACCCCTTTTG 900

QY 948 TGTGTGAA 955

Db 901 TGTGTGAA 908

RESULT 4
ALS55777
LOCUS
DEFINITION ALS55777 LTI_NFL006.Pl2 Homo sapiens cDNA clone CS0DK002YJ18 5

ACCESSION
VERSION ALS55777
KEYWORDS ALS55777.1 GI:12897829

SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
SOURCE
1..881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YJ18"
/clone_1lb="LTI_NFL006.Pl2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 252 a 229 c 225 g 175 t

ORIGIN

Query Match 72.4%; Score 877.8; DB 9; Length 881;
Best Local Similarity 99.8%; Pred. No. 2.7e-184;
Matches 879; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 22 TCACGAGGAGCGGTGGCGCGCGCGCGCGTCCCGGACGCTGTCTGGGACGCGGAG 81

Db 1 TCACGAGGAGCGGTGGCGCGCGCGCGCGTCCCGGACGCTGTCTGGGACGCGGAG 60

QY 82 CCGCGCTGGCCCTTGGAGCGCGCGCGCGCTGATCCCGCGCGCTGTGTCGCCGA 141

Db 61 CCGCGCTGGCCCTTGGAGCGCGCGCGCGCTGATCCCGCGCGCTGTGTCGCCGA 120

QY 142 GTGCGGTGTCCTTGGACCATTCGCCGACGCGCTTCAATTGACGCCGCGAGAGCTCAAGTCC 201

Db 121 GTGCGGTGTCCTTGGACCATTCGCCGACGCGCTTCAATTGACGCCGCGAGAGCTCAAGTCC 180

QY 202 GACCCCGCGCGCGCGCGCGCTTCCAGAGGTTCAATGCTGCTCTGACTGCAAGCGC 261

Db 181 GACCCCGCGCGCGCGCGCTTCCAGAGGTTCAATGCTGCTCTGACTGCAAGCGC 240

QY 262 CAATTACAGCAAGGCTGGAAGCTTGAAGCGCACTTGTGCAACGACACGCGGAGAGAC 321

Db 241 CAATTACAGCAAGGCTGGAAGCTTGAAGCGCACTTGTGCAACGACACGCGGAGAGAC 300

QY 322 ATTTGTTGTGACTATGAGAGGCTGTGCAAGCGCTTCAATGAGGACTACATCTGACCG 381

Db 301 ATTTGTTGTGACTATGAGAGGCTGTGCAAGCGCTTCAATGAGGACTACATCTGACCG 360

QY 382 CCACATTTGACTACACAGAGAGAAAGCGTTTGTGTGACGACATGCTGTGATCA 441

Db 361 CCACATTTGACTACACAGAGAGAAAGCGTTTGTGTGACGACATGCTGTGATCA 420

QY 442 AAAATTCACACAAATGCAAACTTGAAGAAACATTTGAAACGCAACATGAAATCACA 501

Db 421 AAAATTCACACAAATGCAAACTTGAAGAAACATTTGAAACGCAACATGAAATCACA 480

QY 502 AAAACATATATATATGCACTTTTGAAGACTGTAAGAACACTTTAGAACATCAGACCT 561

Db 481 AAAACATATATATATGCACTTTTGAAGACTGTAAGAACACTTTAGAACATCAGACCT 540

QY 562 GAAATTCATGCTGCGCGCATATACATGAAACGCTTTCATAGTATGCCAGAGAGATG 621

Db 541 GAAATTCATGCTGCGCGCATATACATGAAACGCTTTCATAGTATGCCAGAGAGATG 600

QY 622 TGGGAAACACTTTGATACCCAGCAAGCTGAAACGACATGCCAAGCCCGAGAGGCTA 681

Db 601 TGGGAAACACTTTGATACCCAGCAAGCTGAAACGACATGCCAAGCCCGAGAGGCTA 660

QY 682 TGTATGTCAAAAGGATGTTCTTGTGTGCAAAAACATGACGACACTTGAACATGT 741

Db 661 TGTATGTCAAAAGGATGTTCTTGTGTGCAAAAACATGACGACACTTGAACATGT 720

QY 742 GAGAGAAACCATTAAGGGAATATATGTGAAGTATGCGGGAATTTAAACGCA 801

Db 721 GAGAGAAACCATTAAGGGAATATATGTGAAGTATGCGGGAATTTAAACGCA 780

QY 802 AGATTACCTTAAGCAACATGAAACTCATGCCCCAGAAAGGATGTATGTCGTGCC 861

Db 781 AGATTACCTTAAGCAACATGAAACTCATGCCCCAGAAAGGATGTATGTCGTGCC 840

QY 862 AAGAGAGGCTGTGGAAGAACTTACTACTGTGTTTATC 902

Db 841 AAGAGAGGCTGTGGAAGAACTTACTACTGTGTTTATC 881

RESULT 5
BMS62682
LOCUS BMS62682 1059 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6588903 NIH_MGC_98 Homo sapiens cDNA IMAGE:5478431

ACCESSION
VERSION BMS62682
KEYWORDS BMS62682.1 GI:1808978

SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LCM1998 row: a column: 24
 High quality sequence stop: 717.
 Location/Qualifiers

FEATURES

source

1. 1059
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5478431"
 /clone_lib="NIH_MGC_98"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library." 1 others

BASE COUNT 305 a 274 c 270 g 209 t 1 others
 ORIGIN

Query Match 71.8%; Score 870.6; DB 13; Length 1059;
 Best Local Similarity 97.3%; Pred. No. 1.le-182;
 Matches 938; Conservative 0; Mismatches 20; Indels 6; Gaps 5;

35 GTGGGCGGGGGCGGGGTCGCGGACAGTGTCTGGCAGCGGCGGCGCTGGCCCT 94
 Db 1 GTGGGCGGGGGCGGGGTCGCGGACAGTGTCTGGCAGCGGCGGCGCTGGCCCT 60
 95 GGGCTTGGAGGCGCGCGCCGTGGATCCGCGCGCGCTGCGCGAGTCGGTGTGTC 154
 Db 61 GGGCTTGGAGGCGCGCGCGCCGTGGATCCGCGCGCGCTGCGCGAGTCGGTGTGTC 120
 155 TTGACCATGCGCGGCGCGCTTTCATTCGACCGCGGAGAGTCAGTCCGAGCCCGCGCGC 214
 Db 121 TTGACCATGCGCGGCGCGCTTTCATTCGACCGCGGAGAGTCAGTCCGAGCCCGCGCGC 180
 215 CCGGCGCTTCCAGAGAGGTTCACTGCTCTTCCCTGACTGACAGCGCCAAATTACAGCAAA 274
 Db 181 CCGGCGCTTCCAGAGAGGTTCACTGCTCTTCCCTGACTGACAGCGCCAAATTACAGCAAA 240
 275 GCCTGGAAGCTTGAGCGGACCTGTCGAAGCAGCAGGGGAGAGACCATTTGTTGTGAC 334
 Db 241 GCCTGGAAGCTTGAGCGGACCTGTCGAAGCAGCAGGGGAGAGACCATTTGTTGTGAC 300
 335 TATGAAGGTTGGGCAAGGCTTTCATCAGGAGTACCATCTGAGCCGCACATTTGACT 394
 Db 301 TATGAAGGTTGGGCAAGGCTTTCATCAGGAGTACCATCTGAGCCGCACATTTGACT 360
 395 CACACAGCAAAAGCGCGTTGTTGTGCAGCAGCTGGCTGTGATCAAAATTCAGCACA 454
 Db 361 CACACAGCAAAAGCGCGTTGTTGTGCAGCAGCTGGCTGTGATCAAAATTCAGCACA 420
 455 AATCAACTTGAAGAAATTTGAAACGCAACATGAATAAACAACAAACAAATATATA 514
 Db 421 AATCAACTTGAAGAAATTTGAAACGCAACATGAATAAACAACAAACAAATATATA 480
 515 TGCAGTTTGAAGACTGTAGAAGACCTTTAAGAAATCAGAGGCTGAATAATCCATGAC 574
 Db 481 TGCAGTTTGAAGACTGTAGAAGACCTTTAAGAAATCAGAGGCTGAATAATCCATGAC 540
 575 TGCACACATACCAATGAACCTCTATTCAAGTGTACCCAGAGAGAGATGTGGAAACCTTT 634
 Db 541 TGCACACATACCAATGAACCTCTATTCAAGTGTACCCAGAGAGAGATGTGGAAACCTTT 600
 635 GCATCACCAGCAAGCTGAAGACATGCCAAGGCCACGAGGGCTATGTATGTCAAAA 694

|||||
 Db 601 GCATCACCAGCAAGCTGAAGACATGCCAAGGCCACGAGGGCTATGTATGTCAAAA 660
 695 GGAATGTTCCCTTTGTGGCAAAAACATGAGACGAACTTCTGAACATGTGAGAAACCCAT 754
 Db 661 GGAATGTTCCCTTTGTGGCAAAAACATGAGACGAACTTCTGAACATGTGAGAAACCCAT 720
 755 AAAGAGAAATACATGTGAGAGTATGCCGGAACAACTTAAAGCAAAATATACCTTAAG 814
 Db 721 AAAGAGAAATACATGTGAGAGTATGCCGGAACAACTTAAAGCAAAATATACCTTAAG 780
 815 CAACACATGAACAACTCATG-CCCGAAGAAAGGATGTATGTGCTGTCGAAGAA-GGCT 872
 Db 781 CAACACATGAACAACTCATGCCCCCAGAAAGGATGTATGTGCTGTCGAAGAAAGGCT 840
 873 GTGGAGAACCTATACCTGTTTATCTCCAAAGCCATACCTCT-CTTCCATGAG 931
 Db 841 GTGGAGAACCTATACCTGTTTATCTCCAAAGCCATACCTCTCTCTTCATGAG 900
 932 GAAAGCGCCCTTTGTG-TGTGAACATGCTG--CTGTGCAAAACATTTGCCAATGAAA 988
 Db 901 GAAGCCCGCCCTTTGTGGGGTGAACATGCTGGGCTTGGGCAAAACATTTGCCAAGGAA 960
 989 CAAA 992
 Db 961 AAAA 964

RESULT 6
 AL574318/c 937 bp mRNA linear EST 16-FEB-2001
 LOCUS AL574318.LTI_NFL006.PL2 Homo sapiens cDNA clone CS001057YN01 3
 DEFINITION prime, mRNA sequence.
 ACCESSION AL574318
 VERSION AL574318.1 GI:12934410
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 937)
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polyes,D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1. 937
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS001057YN01"
 /clone_lib="LTI_NFL006.PL2"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email: fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>

BASE COUNT 209 a 198 c 204 g 321 t 5 others
 ORIGIN

Query Match 71.1%; Score 862.8; DB 9; Length 937;
 Best Local Similarity 99.0%; Pred. No. 5.7e-181;
 Matches 875; Conservative 3; Mismatches 5; Indels 1; Gaps 1;
 330 GTACTATGAAGGTTGTGGCAAGGCTTCATCAGAGCTACCATCTGAGCGCCGACATTC 389

```

|||||
Db 937 GTRACTATGMAAGGTGGGCAAGGCCCTTTCATCGAGGACCTACCTTGGAGCCGCCCATTC 878
|||
Qy 390 TCACTACACAGAGAGAAAAAGCGTTTGTGTCACGCCACTGCTGTATCAAAATTC 449
|||
Db 877 TTACTCACACAGGAGAAAAAGCGTTTGTGTC -GCCAATGCTGTATCAAAATTC 819
|||
Qy 450 ACACAAATCAAACTTGAAAGAACATTTTACGCCAAACATGAAATCAACAAAACAT 509
|||
Db 818 ACACAAATCAAACTTGAAAGAACATTTTACGCCAAACATGAAATCAACAAAACAT 759
|||
Qy 510 ATATTCGAGTTTGGAGACTGTAAAGAACCTTTAAGAAACATCAGAGCTGAAATTC 569
|||
Db 758 ATATTCGAGTTTGGAGACTGTAAAGAACCTTTAAGAAACATCAGAGCTGAAATTC 699
|||
Qy 570 ATCATGCGCAGCATACCAATGAACTTATTCAGAGTACCCAGGAAGATGTGGGAAC 629
|||
Db 698 ATCATGCGCAGCATACCAATGAACTTATTCAGAGTACCCAGGAAGATGTGGGAAC 639
|||
Qy 630 ACTTTGATCAACCCAGAGCTGAACGACATGCCAAGGCCACGAGGCTATGTATTC 689
|||
Db 638 ACTTTGATCAACCCAGAGCTGAACGACATGCCAAGGCCACGAGGCTATGTATTC 579
|||
Qy 690 AAAAGAGTGTCTCTTGTGGCAAAACATGAGAGAACCTTGAACATGTGAGAGAA 749
|||
Db 578 AAAAGAGTGTCTCTTGTGGCAAAACATGAGAGAACCTTGAACATGTGAGAGAA 519
|||
Qy 750 CCCATTAAGAGGAATACTATGTAGAGTATCCGGAAGAACATTTAAAGCCAAAGATTACC 809
|||
Db 518 CCCATTAAGAGGAATACTATGTAGAGTATCCGGAAGAACATTTAAAGCCAAAGATTACC 459
|||
Qy 810 TTAAGCAACATGCAAAATCATGCCCCAGAAAGAGATGTATGCTGCTCCAGAGAA 869
|||
Db 458 TTAAGCAACATGCAAAATCATGCCCCAGAAAGAGATGTATGCTGCTCCAGAGAA 399
|||
Qy 870 GGTGTGGAAGAACTATCTACTGTGTTTAAATCCAAAGCATCTCTCTCTTCATG 929
|||
Db 398 GGTGTGGAAGAACTATCTACTGTGTTTAAATCCAAAGCATCTCTCTCTTCATG 339
|||
Qy 930 AGGAAAGCCGCTTTTGTGTGAACATCTGCTGTGGCAAAACATTTGCAATGAAG 989
|||
Db 338 AGGAAAGCCGCTTTTGTGTGAACATCTGCTGTGGCAAAACATTTGCAATGAAG 279
|||
Qy 990 AAAGTCTCACTAGGATCTCTTGTACATGATCTGCAAGAAAGAAATGAAGCTCAAG 1049
|||
Db 278 AAAGTCTCACTAGGATCTCTTGTACATGATCTGCAAGAAAGAAATGAAGCTCAAG 219
|||
Qy 1050 TCAAAATAATCTGTAAGAAAGAGATTTGGCTCTCATCTCACTGATATATCTCCCA 1109
|||
Db 218 TCAAAATAATCTGTAAGAAAGAGATTTGGCTCTCATCTCACTGATATATCTCCCA 159
|||
Qy 1110 AAAGAAACAGAGGCAAGGCTTATCTTTGTGCAAAAGAGAGTACCCAACTGTGG 1169
|||
Db 158 AAAGAAACAGAGGCAAGGCTTATCTTTGTGCAAAAGAGAGTACCCAACTGTGG 99
|||
Qy 1170 AAAGAAAGATGCTCTGACAGATTTGCACTACTTACCTTGGCTAA 1213
|||
Db 98 AAAGAAAGATGCTCTGACAGATTTGCACTACTTACCTTGGCTAA 55
|||
RESULT 7
BO650498 930 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8207703 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283123
DEFINITION 5', mRNA sequence.
ACCESSION BO650498
VERSION BO650498.1 GI:21774670
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 930)

```

```

AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: L10M2478 row: 3 column: 20
High quality sequence stop: 636.
FEATURES
source
1. .930
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="6283123"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pORF7; Site: 1; XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(C). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the Laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 316 a 212 c 200 g 201 t 1 others
ORIGIN
Query Match 70.9%; Score 860.4; DB 14; Length 930;
Best Local Similarity 97.0%; Pred. No. 1, 9e-180;
Matches 899; Conservative 0; Mismatches 22; Indels 6; Gaps 2;
Qy 241 CTCCTCCCTGACTGACGCGCCCAATTAACAGAAAGCTGGAACCTTGAGCGCACCTG 300
|||
Db 1 CTCCTCCCTGACTGACGCGCCCAATTAACAGAAAGCTGGAACCTTGAGCGCACCTG 60
|||
Qy 301 CAAGCAGCGGGGAGAGACCATTTGTTGTGATGAGAGGTGTGGCAAGCCCTTCAT 360
|||
Db 61 CAAGCAGCGGGGAGAGACCATTTGTTGTGATGAGAGGTGTGGCAAGCCCTTCAT 120
|||
Qy 361 CAGGACTACATCTGAGCCGCCCACTTGACTGACACAGAGAGAAAGCCGTTGTTG 420
|||
Db 121 CAGGACTACATCTGAGCCGCCCACTTGACTGACACAGAGAGAAAGCCGTTGTTG 180
|||
Qy 421 TCGAGCCTGCTGTGATCAAAATTTCAACACAAATCAACTTAAGAAACATTTGA 480
|||
Db 181 TCGAGCCTGCTGTGATCAAAATTTCAACACAAATCAACTTAAGAAACATTTGA 240
|||
Qy 481 ACGCAACATGAAATTCACAAACAAATATATATGAGTGTGAAGACTGTAAAGAGC 540
|||
Db 241 ACGCAACATGAAATTCACAAACAAATATATATGAGTGTGAAGACTGTAAAGAGC 300
|||
Qy 541 CTTTAAAGAACTTCAGAGCTGGAATTCATCACTGCGACATACCAATGAACCTTAT 600
|||
Db 301 CTTTAAAGAACTTCAGAGCTGGAATTCATCACTGCGACATACCAATGAACCTTAT 360
|||
Qy 601 CAAGTGTACCCAGGAAGATGTGGGAAACACTTGTGCTACCCAGCAAGCTGAAGACA 660
|||
Db 361 CAAGTGTACCCAGGAAGATGTGGGAAACACTTGTGCTACCCAGCAAGCTGAAGACA 420
|||
Qy 661 TGGCAAGGCCCGAGAGGCTATGTATGTCAAAAGATGTCTCTTGTGGCAAAACATG 720
|||
Db 421 TGGCAAGGCCCGAGAGGCTATGTATGTCAAAAGATGTCTCTTGTGGCAAAACATG 480
|||
Qy 721 GACGGAATCTTGAAACATGTGAGAGAAACCATTAAGAGGAATCTATGCAAGTATG 780
|||
Db 481 GACGGAATCTTGAAACATGTGAGAGAAACCATTAAGAGGAATCTATGCAAGTATG 540

```


BASE COUNT 250 a 232 c 230 g 177 t 4 others

Query Match 69.2%; Score 838.8; DB 9; Length 893;
Best Local Similarity 99.4%; Pred. No. 1.2e-175;
Matches 840; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GTGCGGCGCGCGGAGGTTCAAGAGAGCGGTGGCGCGCGCGCGGTCCTCCGCG 60
DB 2 GTGCGGCGCGCGGAGGTTCAAGAGAGCGGTGGCGCGCGCGCGGTCCTCCGCG 61
OY 61 AGCTGTCTGCGGACGTGGCAGCGCGCGCTGGCGCTTGAGAGCGCGCGCGCTTGA 120
DB 62 AGCTGTCTGCGGACGTGGCAGCGCGCGCTGGCGCTTGAGAGCGCGCGCGCTTGA 121
OY 121 TCGCGCGCGCGGTGGCGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
DB 122 TCGCGCGCGCGGTGGCGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 181
OY 181 AGCGGCGGAGAGCTCAGTCCGACCGCGCGCGCGCGCTTCCAGAGAGTTTCATCTG 240
DB 182 AGCGGCGGAGAGCTCAGTCCGACCGCGCGCGCGCGCTTCCAGAGAGTTTCATCTG 241
OY 241 CTCCTTCCTGAGCTGAGCGCAATTACAGCAAGCGTGGAGAGTTCAGCGACCTGG 300
DB 242 CTCCTTCCTGAGCTGAGCGCAATTACAGCAAGCGTGGAGAGTTCAGCGACCTGG 301
OY 301 CAAGCACAGCGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGGCGAGGCTTCAT 360
DB 302 CAAGCACAGCGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGGCGAGGCTTCAT 361
OY 361 CAGGACTACATCTGAGCGCGCAATCTGACTACACAGAGAGAGAGAGCGTTGTTG 420
DB 362 CAGGACTACATCTGAGCGCGCAATCTGACTACACAGAGAGAGAGAGCGTTGTTG 421
OY 421 TGCAGGCACTGCTGTGATCAAAAATTCACAAAATCAAACTTGAAGAACATTTTGA 480
DB 422 TGCAGGCACTGCTGTGATCAAAAATTCACAAAATCAAACTTGAAGAACATTTTGA 481
OY 481 AGCGAAGCATGAAATTCACAAAATCATATATATGAGAGTGGAGAGTGAAGAGAC 540
DB 482 AGCGAAGCATGAAATTCACAAAATCATATATATGAGAGTGGAGAGTGAAGAGAC 541
OY 541 CTTTAAAGAAACATCAGCAGCTGAAATTCATCAGTGCAGCATACCAATGAACCTGTAT 600
DB 542 CTTTAAAGAAACATCAGCAGCTGAAATTCATCAGTGCAGCATACCAATGAACCTGTAT 601
OY 601 CAGGTGTACCCAGGAGAGTGGGGAACATTTGCATCACCGACAGCTGAAGAGACA 660
DB 602 CAGGTGTACCCAGGAGAGTGGGGAACATTTGCATCACCGACAGCTGAAGAGACA 661
OY 661 TGCAGAGGCCACGAGGCGTATGATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 720
DB 662 TGCAGAGGCCACGAGGCGTATGATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 721
OY 721 GAGGAGACTTTGAAACATGTGAGAGAAACCATTAAGAGAGAAATATATGTAAGTATG 780
DB 722 GAGGAGACTTTGAAACATGTGAGAGAAACCATTAAGAGAGAAATATATGTAAGTATG 781
OY 781 CCGGAAACATTTAAACGAAAGATTCCTTAAGCAACACATGAAACCTCATGCCGACA 840
DB 782 CCGGAAACATTTAAACGAAAGATTCCTTAAGCAACACATGAAACCTCATGCCGACA 841
OY 841 AAGGG 845
DB 842 AANGG 846

RESULT 11
LOCUS B0644275 902 bp mRNA linear EST 15-JUL-2002
AGENCY/RT 8355851 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285421
DEFINITION 5', mRNA sequence.

ACCESSION B0644275
VERSION B0644275.1 GI:21768447
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 902)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@briemail.nih.gov
Tissue Procurement: CGAP (Stanford)

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2484 row: j column: 14
High quality sequence stop: 642.

FEATURES
source location/qualifiers
1..902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:6285421"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pORF7; Site: 1; XhoI: Site: 2;
EcoRI: cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGGACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 301 a 208 c 197 g 194 t 2 others

Query Match 68.9%; Score 836; DB 14; Length 902;
Best Local Similarity 98.3%; Pred. No. 4.9e-175;
Matches 875; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

OY 240 GCTCTTCCCTGAGTGGAGCGCAATTACAGCAAGCTGGAGCTTGAAGCGCACCTGT 295
DB 11 GCTCTTCCCTGAGTGGAGCGCAATTACAGCAAGCTGGAGCTTGAAGCGCACCTGT 70
OY 300 GCAAGCACAGGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGTGGCAGGCTTCA 359
DB 11 GCTCTTCCCTGAGTGGAGCGCAATTACAGCAAGCTGGAGCTTGAAGCGCACCTGT 70
OY 71 GC-AMCAGACGGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGTGGCAGGCTTCA 129
DB 360 TCAAGGACTACATCTGAGCGCGCAATTCGATCAGCAGAGAGAGAGAGAGAGAGAG 419
DB 130 TCAAGGACTACATCTGAGCGCGCAATTCGATCAGCAGAGAGAGAGAGAGAGAGAG 189
OY 420 GTGACGCACTGCTGTGATCAAAAATTCACAGCAAAAATCAAACTTGAAGAACATTTTG 479
DB 190 GTGACGCACTGCTGTGATCAAAAATTCACAGCAAAAATCAAACTTGAAGAACATTTTG 479
OY 480 AACGCAACATGAAATCAAGCAAAAATCATATATGAGAGAGAGAGAGAGAGAGAGAG 539
DB 250 AACGCAACATGAAATCAAGCAAAAATCATATATGAGAGAGAGAGAGAGAGAGAGAG 309
OY 540 CTTTAAAGAAACATCAGCAGCTGAAATTCATCAGTCCAGCATACCAATGAACCTGTAT 599
DB 310 CTTTAAAGAAACATCAGCAGCTGAAATTCATCAGTCCAGCATACCAATGAACCTGTAT 369
OY 600 TCAAGGTACCCAGGAGAGATGTGGAGAACATTTGGATACCCAGAGAACCTGAAGAGAC 659
DB 370 TCAAGGTACCCAGGAGAGATGTGGAGAACATTTGGATACCCAGAGAACCTGAAGAGAC 429


```

OY 660 ATCCGAGGCCGACGAGGGCTATGTATGTCAGAAAAAGGATGCTTCTTGGGCAAAAAACAT 719
    |||||||
Db 430 ATGCCAAGGCCCGACGAGGGCTATGTATGTCAGAAAAAGGATGCTTCTTGGGCAAAAAACAT 489
OY 720 GGACGGAACCTTCTGGAACATGTGAGAGAAACCCATNAAGAGAAATACTATGTGAAGTAT 779
    |||||||
Db 490 GGACGGAACCTTCTGGAACATGTGAGAGAAACCCATNAAGAGAAATACTATGTGAAGTAT 549
OY 780 GCGGGAACAACTTTAAAGCAAAATGTAAGTAAAGCAACATGAAACTGATGCGCCAG 839
    |||||||
Db 550 GCGGGAACAACTTTAAAGCAAAATGTAAGTAAAGCAACATGAAACTGATGCGCCAG 609
OY 840 AAAGGATGTATGTGCTGTCGTCAGAGAGAGGCTGTGAGAAACCTTACTAGTGTGTTA 899
    |||||||
Db 610 AAAGGATGTATGTGCTGTCGTCAGAGAGAGGCTGTGAGAAACCTTACTAGTGTGTTA 669
OY 900 ATCTCCAAAGCCATATCTCTCTCCATGAGAAAGCCGCTTTGTGTGTGAACATG 959
    |||||||
Db 670 ATCTCCAAAGCCATATCTCTCTCCATGAGAGAAAGCCGCTTTGTGTGTGAACATG 729
OY 960 CTGGCTGTGGCAAAACATTTSCAATGAAGCAAGTCTACTAGGATGCTGTGTGATG 1019
    |||||||
Db 730 CTGGCTGTGGCAAAACATTTSCAATGAAGCAAGTCTACTAGGATGCTGTGTGATG 789
OY 1020 ATCTGACAGAGAAATGAAGCTCAAGCAAAAAATCTGTAAGAAAGGAGTTGG 1079
    |||||||
Db 790 ATCTGACAGAGAAAG-AAATGAAGCTCCAGTCCAAAAATCTGTGAAAAAGGAGTTGG 848
OY 1080 CCTCTCATCTCAGTGGATATAT-CCCTCCCAAAAGGAACAGGCGCAGG 1128
    |||||||
Db 849 CCTCTCATCTCAGTGGATATATCCCTCCCAAAAGGAACAGGCGCAGG 898

```

```

RESULT 12
AL578193/c 933 bp mRNA linear EST 16-FEB-2001
LOCUS AL578193 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK002YJ18 3
DEFINITION prime, mRNA sequence.
ACCESSION AL578193
VERSION AL578193.1 GI:12942034
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 933)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
            Genoscope - Centre National de Sequencage
            Bp 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
    source
        1. 933
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DK002YJ18"
            /clone_id="LTI_NFL006_PL2"
            /tissue_type="placenta"
            /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed by
            Life Technologies. Contact: Feng Liang Life Technologies,
            Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
            Email : fliang@life-tech.com URL :
            http://fulllength.lifetech.com"

```

```

BASE COUNT 202 a 192 c 207 g 326 t 6 others
ORIGIN

```

```

Query Match 68.4% Score 830.2; DB 9; Length 933;
Best Local Similarity 98.5% Pred. No. 9,5e-174;
Matches 854; Conservative 4; Mismatches 7; Indels 2; Gaps 2;

OY 348 GCAAGGCTTCATCAGGAGACTACATCTGAGCCGCCACATTCGTGACACAGAGGAAA 407
    |||||||
Db 933 GCAAGGCTTCATCA-GGAGTACCATCTGAGAGCCGCCACATTCGTGACACAGAGGAAA 875
OY 408 AGCCGTTGTTGTGACCCACCTGGCTGTGTCAAAAATTCAACACAAAATCAAACTTGA 467
    |||||||
Db 874 AGCCGTTGTTGTGACCCACCTGGCTGTGTCAAAAATTCAACACAAAATCAAACTTGA 815
OY 468 AGAATCTTTGAGCAAGCAACATGAAATCAACAAAAACATATATATGACGTTTGAAG 527
    |||||||
Db 814 AGAATCTTTGAGCAAGCAACATGAAATCAACAAAAACATATATATGACGTTTGAAG 755
OY 528 ACTGTGAAGAGCCTTTAAGAAACATGACAGCTGAAATACCTACAGTCCAGCATACA 587
    |||||||
Db 754 ACTGTGAAGAGCCTTTAAGAAACATGACAGCTGAAATACCTACAGTCCAGCATACA 695
OY 588 ATGAACCTCTATTCAAGTGTACCGAGAGGATGTGGAAACACTTTCATGACCCAGCA 647
    |||||||
Db 694 ATGAACCTCTATTCAAGTGTACCGAGAGGATGTGGAAACACTTTCATGACCCAGCA 635
OY 648 ATGGAAGACAGATGCCAAGGCCACGAGGCTATGTTATGTCAAAAGATGTTCTTTG 707
    |||||||
Db 634 ATGGAAGACAGATGCCAAGGCCACGAGGCTATGTTATGTCAAAAGATGTTCTTTG 575
OY 708 TGGCAAAAACATGACGAGACCTTGTGAACATGTGAGAGAAACCCATGAAGAGAAATAC 767
    |||||||
Db 574 TGGCAAAAACATGACGAGACCTTGTGAACATGTGAGAGAAACCCATGAAGAGAAATAC 515
OY 768 TATGTGAAATATGCGGGAACATTTAAAGCAAAAGTTCCTTAAGCAACATGAAAA 827
    |||||||
Db 514 TATGTGAAATATGCGGGAACATTTAAAGCAAAAGTTCCTTAAGCAACATGAAAA 455
OY 828 CTCATGCCCCAGAAAGGAGATGATGCGCTGTCCAGAGAGAGCTGTGAAAGAACTATA 887
    |||||||
Db 454 CTCATGCCCCAGAAAGGAGATGATGCGCTGTCCAGAGAGAGCTGTGAAAGAACTATA 395
OY 888 CTACGAGTGTATATCTCCAAAGCAATCCCTCCCTCCATGAGAGAAAGCGGCTTTTG 947
    |||||||
Db 394 CAACGTGTTTATATCTCCAAAGCAATCTCTCTCTCCATGAGAGAAAGCGGCTTTTG 335
OY 948 TGTGTGAACATCTGCTGTGTGCAAAACATTTGCAATGAAGCAAGTCTCCTAGGCAATG 1007
    |||||||
Db 334 TGTGTGAACATCTGCTGTGTGCAAAACATTTGCAATGAAGCAAGTCTCCTAGGCAATG 275
OY 1008 CTGTGTATACATGATCTGCAAGAAAGAAATGAAGCTCAAGTCAAAAAATCTCGTGA 1067
    |||||||
Db 274 CTGTGTATACATGATCTGCAAGAAAGAAATGAAGCTCAAGTCAAAAAATCTCGTGA 215
OY 1068 AACGAGTTTGGCTCTCATCTCAAGTGAATATATCCCTGCCAAAGAGAAACAGGCAAG 1127
    |||||||
Db 214 AACGAGTTTGGCTCTCATCTCAAGTGAATATATCCCTGCCAAAGAGAGCAAGGCAAG 155
OY 1128 GCTTATCTTTGTGTCAAAACGAGAGTCAACCACTGTGGAAGCAAGAGCTCTGCA 1187
    |||||||
Db 154 GCTTATCTTTGTGTCAAAACGAGAGTCAACCACTGTGGAAGCAAGAGCTCTGCA 95
OY 1188 CAGTTGCA-GTACTTACCCTTGCTAA 1213
    |||||||
Db 94 CAGTTGCAAGCACTTACCCTTGCTAA 68

RESULT 13
BO278693 1089 bp mRNA linear EST 07-MAY-2002
LOCUS BO278693
DEFINITION AGENCOURT_7049694 NIH_MGC_107 Homo sapiens cDNA clone IMAGE:5805333
ACCESSION BO278693
VERSION BO278693.1 GI:20488901

```


KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 1089)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM204 row: f column: 22
High quality sequence stop: 610.
Location/Qualifiers
1. 1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5805333"
/clone_lib="NIH-MGC_107"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library."

BASE COUNT 296 a 289 c 283 g 221 t
ORIGIN

Query Match 68.3%; Score 829; DB 14; Length 1089;
Best Local Similarity 94.6%; Pred. NO. 1.8e-173;
Matches 926; Conservative 0; Mismatches 40; Indels 13; Gaps 6;

0Y 11 CCGCGGAAGGTTCAAGCAAGGAGCGGTGGCGCGCGCGGTCCCGGACGCTGTCTCG 70
2 CCGCGGAAGGTTCAAGCAAGGAGCGGTGGCGCGCGCGGTCCCGGACGCTGTCTCG 61
71 GCACGAGGAGCGCGCGGTGGCGCGCGGTGGCGCGCGCGGTCCCGGACGCTGTCTCG 130
62 GCACGAGGAGCGCGCGGTGGCGCGCGGTGGCGCGCGCGGTCCCGGACGCTGTCTCG 121
0Y 131 GTGTGTGCGCGAGTGTGTCTCTTCAACATCGCGACGCTTCATTCGACCGCGCGAG 190
122 GTGTGTGCGCGAGTGTGTCTCTTCAACATCGCGACGCTTCATTCGACCGCGCGAG 181
0Y 191 AGCTGAGTCCGAGACCGCGCGCGCGCGGTCCCGGAGGTCAATCTGCTCTTCCCT 250
182 AGCTGAGTCCGAGACCGCGCGCGCGCGGTCCCGGAGGTCAATCTGCTCTTCCCT 241
0Y 251 GACTGACGCGCAATTTACAGCAAAAGCTGGAAGCTTGACGCGACCTGTGCAAGCACAG 310
242 GACTGACGCGCAATTTACAGCAAAAGCTGGAAGCTTGACGCGACCTGTGCAAGCACAG 301
0Y 311 GGGGAGAGACATTTGTTGTGACTATGAGAGGTGTGCAAGGCTTCATCAAGGAGCTAC 370
302 GGGGAGAGACATTTGTTGTGACTATGAGAGGTGTGCAAGGCTTCATCAAGGAGCTAC 361
0Y 371 CATCTGAGCGCGCACATTTCTGACTACACAGAGAAAGCCGTTTGTGTGACGACCT 430
362 CATCTGAGCGCGCACATTTCTGACTACACAGAGAAAGCCGTTTGTGTGACGACCT 421
0Y 431 GGCTGTGATCAAAATTTCAACGAAATCAAACTGAAGAAATTTTGAAGCAAAAT 490
|||||

Db 422 GGCTGTGATCAAAATTTCAACGAAATCAAACTTGAAGAAATTTTGAAGCAAAAT 481
0Y 491 GAAATCAACAAATAATATATATGAGTTTGAAGACTGTAAAGACCTTTAAGAAA 550
482 GAAATCAACAAATAATATATATGAGTTTGAAGACTGTAAAGACCTTTAAGAAA 541
0Y 551 CATCAACAGCTAAATATCATGAGCGACATGCAATGCAATGCAATGCAATGCAAT 610
542 CATCAACAGCTAAATATCATGAGCGACATGCAATGCAATGCAATGCAATGCAAT 601
0Y 611 CAGGAAGATGTGGGAAACACTTTCATCAACCCAGCAAGCTGAAAGAGATGCCAGGCC 670
602 CAGGAAGATGTGGGAAACACTTTCATCAACCCAGCAAGCTGAAAGAGATGCCAGGCC 661
0Y 671 CAGGAGGCTATGTATGTCATAAAGAGAGTCTTGTGCAAAAATGAGAGAGACT 730
662 CAGGAGGCTATGTATGTCATAAAGAGAGTCTTGTGCAAAAATGAGAGAGACT 721
0Y 731 CTGAACATGTAGAGAAACCATTAAGAGCAATTAATGTGTAAGTATGCCGAAACA 790
722 CTGAACATGTAGAGAAACCATTAAGAGCAATTAATGTGTAAGTATGCCGAAACA 781
0Y 791 TTAAAGCAAGATTTACTTATAGCAACATGAATCAAT--GCCCGAAGAGGATG 848
782 TTAAAGCAAGATTTACTTATAGCAACATGAATCAAT--GCCCGAAGAGGATG 841
Db 842 TATGTCCCTGTGCCCAAAAGAGGTGTGCAAAAACCTTTCACGTGGGTTATCT 901
0Y 849 TATGTCC--TGTCCAAAGAGGCTGTGG--AAGAACCTTACTACTGT--GTTAAATCT 903
842 TATGTCCCTGTGCCCAAAAGAGGTGTGCAAAAACCTTTCACGTGGGTTATCT 901
0Y 904 CCAAGCCTATC---CTCTCTTCATGAGAGAGAGGCGCTTT---GTGTGACAA 957
902 CCAAGCCTATCCTCTCTCTTCCTGTAAGAGAGAGGCGCTTTTGGGGGGGGGAA 961
0Y 958 TGCTGGCTGTGCAAAACA 976
962 TGCTGGCTGTGCGCAAAA 980
Db

RESULT 14
LOCUS BG830080 927 bp mRNA linear EST 22-MAY-2001
DEFINITION 602764826f1 NIH-MGC_42 Homo sapiens cDNA clone IMAGE:4906710 5',
mRNA sequence.
ACCESSION BG830080
VERSION BG830080.1 GI:14177667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 927)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1810 row: d column: 07
High quality sequence stop: 880.
Location/Qualifiers
1. 927
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4906710"
/clone_lib="NIH-MGC_42"
/tissue_type="epithelioid carcinoma cell line"

/lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;
 Site_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library. I"

BASE COUNT 264 a 244 c 229 g 190 t
 ORIGIN

Query Match 68.2%; Score 827.4; DB 12; Length 927;

Best Local Similarity 98.6%; Pred. No. 4e-173;
 Matches 908; Conservative 0; Mismatches 6; Indels 7; Gaps 7;

```

OY 38 GCGCGGGCGCGCGCGTTCGCCGACGCTCTCGGACGTGGCAGCGCGCTTGGCCCTGGG 97
DB 1 GCGCGGGCGCGCGCGTTCGCCGACGCTCTCGGACGCTGGCAGCGCGCTTGGCCCTGGG 60
OY 98 CTTGGAGGCGCGCGCGCTTGCATCCCGCGCGCTGGTGCAGTGGGTGTCGTCCTTG 157
DB 61 CTTGGAGGCGCGCGCGCTTGCATCCCGCGCGCTGGTGCAGTGGGTGTCGTCCTTG 120
OY 158 ACCATCGCGCGCGCTTGCATCCCGCGCGCTGGTGCAGTGGGTGTCGTCCTTG 217
DB 121 ACCATCGCGCGCGCTTGCATCCCGCGCGCTGGTGCAGTGGGTGTCGTCCTTG 180
OY 218 GCGCTTCCGAGAGGTTTCATCTGCTCTCTCTGCTGACGCGCCCAATTACAGCAAGCC 277
DB 181 GCGCTTCCGAGAGGTTTCATCTGCTCTCTCTGCTGACGCGCCCAATTACAGCAAGCC 240
OY 278 TGGAGGCTTGGAGCGCGCGCTTGCAGCAGCGGGGAGAGCAATTCTTTGTGACTAT 337
DB 241 TGGAGGCTTGGAGCGCGCGCTTGCAGCAGCGGGGAGAGCAATTCTTTGTGACTAT 300
OY 338 GAAGGCTTGGCAAGCGCTTGCATCAGGAGTACCATCTGAGCGCCCATCTTGACTAC 397
DB 301 GAAGGCTTGGCAAGCGCTTGCATCAGGAGTACCATCTGAGCGCCCATCTTGACTAC 360
OY 398 ACAGGAGAAAGCGCTTGTGTGACGACGCGCTGATCAAAATTCAACACAAA 457
DB 361 ACAGGAGAAAGCGCTTGTGTGACGACGCGCTGATCAAAATTCAACACAAA 420
OY 458 TCAAACTTGAAGAAATTTTGAAGCGCAACATGAATAACAAACATATATATGCG 517
DB 421 TCAAACTTGAAGAAATTTTGAAGCGCAACATGAATAACAAACATATATATGCG 480
OY 518 ACTTTTGAAGACTGTAAAGAGCTTTAAAGAAATCAGAGCTGAATAATCATCATGCG 577
DB 481 ACTTTTGAAGACTGTAAAGAGCTTTAAAGAAATCAGAGCTGAATAATCATCATGCG 540
OY 578 CAGCATACCAATGAACCTTATTCAGTGTACCCAGAGAGGATGTGGAAACACTTTGCA 637
DB 541 CAGCATACCAATGAACCTTATTCAGTGTACCCAGAGAGGATGTGGAAACACTTTGCA 600
OY 638 TCAGCAGCAGAGCTGAAAGCGACATGCCAGAGCGCGCTATGATCAAAAAGGA 697
DB 601 TCAGCAGCAGAGCTGAAAGCGACATGCCAGAGCGCGCTATGATCAAAAAGGA 660
OY 698 TGTTCCTTGTGGCAAAACATGAGCAGCACTTCTGAACATGTGAGAAACCCAT-AA 756
DB 661 TGTTCCTTGTGGCAAAACATGAGCAGCACTTCTGAACATGTGAGAAACCCATAAA 720
OY 757 AAGGAATACATGCTG-AAATATGCGGGAAGAA-CATTTAAACAGCAAA-A-TTACTTAA 813
DB 721 AAGGAATACATGCTGAAAGTATGCGGGAAGAAACATTTAAAGCGCAAAATTTACTTAA 780
OY 814 GCACACATGAAAGCTATGCCCGCAGAAAGGATGTATGCTGTCCAGAGAGAGGCT- 872
DB 781 GCACACATGAAAGCTATG-CCGAGAAAGGATGTATGCTGTCCAGAGAGAGGCTG 839

```

```

OY 873 GTGGAGAAACCTATATCTATCTGTTTAAAT-CTCCAAAGCATATCTCTTCATGAG 931
DB 840 GTGGAGAAACCTATATCTATCTGTTTAAATCTCCAAAGCATATCTCTTCATGAG 899
OY 932 GAAAGCGCGCGCTTGTGTGTGT 952
DB 900 GAAAGCGCGCGCTTGTGTGTGT 920

```

RESULT 15

BO648620

LOCUS

DEFINITION BO648620 908 bp mRNA linear EST 15-JUL-2002

AGENCOURT 8415143 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6271890

ACCESSION BO648620

VERSION BO648620.1 GI:21772792

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgi.nci.nih.gov/
 1 (bases 1 to 908)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 f column: 19
 Place: LDCM2449 row: 608.
 High quality sequence stop: 608.
 Location/Qualifiers

FEATURES

source

1..908

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6271890"

/clone_1bp="NIH_MGC_100"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; CDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC
 library."

BASE COUNT

306 a 208 c 196 g 198 t

ORIGIN

Query Match 68.2%; Score 827.2; DB 14; Length 908;

Best Local Similarity 97.4%; Pred. No. 4.e-173;
 Matches 884; Conservative 0; Mismatches 18; Indels 6; Gaps 4;

```

OY 241 CTCCTTCCCTGACGACGCGCAATTACAGCAAGCTTGAGACCTTGACGCGACCTGTG 300
DB 1 CTCCTTCCCTGACGACGCGCAATTACAGCAAGCTTGAGACCTTGACGCGACCTGTG 60
OY 301 CAAGCAGCGGGGGAGACCACTTTGTTGCTAAGAAAGGTGGGCAAGGCTTCAAT 360
DB 61 CAAGCAGCGGGGGAGACCACTTTGTTGCTAAGAAAGGTGGGCAAGGCTTCAAT 120
OY 361 CAGGAGCTACCATCTGACCGCGCACATCTTGACTCACACAGAGGAAAGACCGCTTGTG 420
DB 121 CAGGAGCTACCATCTGACCGCGCACATCTTGACTCACACAGAGGAAAGACCGCTTGTG 180
OY 421 TGCAGCCTACTGGCTGTGATCAAAAATTCACACAAATTCAAAACCTTGAAGAAACATTGTA 480

```

```
||||| 181 TGCAGCCATGCTGTGTCATAAAATTCACACAAAATCAAACTTGAGAGAAACATTTGA 240
OY 481 ACGCAACATGAAAAATACAAAAACAATATATATGCACTTTGAAGACTGTAAAGAC 540
Db 241 ACGCAACATGAAAAATACAAAAACAATATATATGCACTTTGAAGACTGTAAAGAC 300
OY 541 CTTTAAAGAAACATGCGAGCTGAAATTCATCGTGCACGATACCAATGAACCTCTATT 600
Db 301 CTTTAAAGAAACATGCGAGCTGAAATTCATCGTGCACGATACCAATGAACCTCTATT 360
OY 601 CAAGTGTACCCAGGAAGATGTGGAAACACTTTGCATCCAGCAAGCTGAACGACA 660
Db 361 CAAGTGTACCCAGGAAGATGTGGAAACACTTTGCATCCAGCAAGCTGAACGACA 420
OY 661 TGGCAAGGCCCAAGGGCTATGTATGCAAAAAGATGTTCTCTGTGGCAAAAACATG 720
Db 421 TGGCAAGGCCCAAGGGCTATGTATGCAAAAAGATGTTCTCTGTGGCAAAAACATG 480
OY 721 GACGGAACCTTGTGAACATGTGAGAGAAACCATAAAGAGAAATCTATGTGAATATG 780
Db 481 GACGGAACCTTGTGAACATGTGAGAGAAACCATAAAGAGAAATCTATGTGAATATG 540
OY 781 CCGGAAACATTTAAACCAAGATTAACCTTAAGCACACATGAAGAACTATGCCCCAGA 840
Db 541 CCGGAAACATTTAAACCAAGATTAACCTTAAGCACACATGAAGAACTATGCCCCAGA 600
OY 841 AAGGATGTATGTGCTGTCCAGAGAGAGGCTGTGAAGAACCTATACTACTGTGTTAA 900
Db 601 AAGGATGTATGTGCTGTCCAGAGAGAGGCTGTGAAGAACCTATACTACTGTGTTAA 660
OY 901 TCTCCAAAGCCATATCTCTCTCCATGAGGAAAGCCGCTTTGTGTGTGAACAT 958
Db 661 TCTCCAAAGCCATATCTCTCTCCATGAGGAAAGCCGCTTTGTGTGTGAACAT 720
OY 958 GCTGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCACATAAGGATGCTGTGTACAT 1018
Db 721 GCTGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCACATAAGGATGCTGTGTACAT 780
OY 1018 GATCCGTGACAAAGAAATGAGCTCAAAATCTCGTGAAGAAACGAGATTG 1078
Db 781 GATCCGTGACAAAG-AGAAATGAGAGCTCAAAATCTCGTGAAGAAACGAGATTG 839
OY 1078 G-CCTCTCATCTCACTGATGATATATCC--TCCCAAAAGAAACAGGGCAAGCTTATCT 1135
Db 840 GCCCTCTCATCTCACTGATGATATATCCCTCCAAAAGAAAGGCAAGCTTATCT 899
OY 1136 TTGTGTCA 1143
Db 900 TTGGGGCA 907
```

Search completed: February 10, 2003, 17:44:18
Job time : 1907.59 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:45:57 : Search time 40 Seconds
(without alignments)
739.844 Million cell updates/sec

Title: US-09-831-426c-2

Perfect score: 2008
Sequence: 1 MPPAVVAESVSSLTADAF.....ESPNCVEDKMLSTVAVLTLG 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 405691 seqs, 81078759 residues

Total number of hits satisfying chosen parameters: 405691

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2.6/pdata/2/paa/PCIT_NEW_COMB.pep:*
2: /cgn2.6/pdata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2.6/pdata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2.6/pdata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2.6/pdata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2.6/pdata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2.6/pdata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2000	99.6	430	US-09-724-676-93212	Sequence 93212, A
2	2000	99.6	430	US-09-724-676-93223	Sequence 93223, A
3	2000	99.6	430	US-09-724-676-93234	Sequence 93234, A
4	2000	99.6	430	US-09-724-676-93245	Sequence 93245, A
5	2000	99.6	430	US-09-724-676-93256	Sequence 93256, A
6	2000	99.6	430	US-09-724-676-93267	Sequence 93267, A
7	2000	99.6	430	US-09-724-676-93278	Sequence 93278, A
8	2000	99.6	430	US-09-724-676-93289	Sequence 93289, A
9	2000	99.6	430	US-09-724-676-93300	Sequence 93300, A
10	2000	99.6	430	US-09-724-676A-93212	Sequence 93212, A
11	2000	99.6	430	US-09-724-676A-93223	Sequence 93223, A
12	2000	99.6	430	US-09-724-676A-93234	Sequence 93234, A
13	2000	99.6	430	US-09-724-676A-93245	Sequence 93245, A
14	2000	99.6	430	US-09-724-676A-93256	Sequence 93256, A
15	2000	99.6	430	US-09-724-676A-93267	Sequence 93267, A
16	2000	99.6	430	US-09-724-676A-93278	Sequence 93278, A
17	2000	99.6	430	US-09-724-676A-93289	Sequence 93289, A
18	2000	99.6	430	US-09-724-676A-93300	Sequence 93300, A
19	1752	87.3	391	US-09-724-676-93213	Sequence 93213, A
20	1752	87.3	391	US-09-724-676-93214	Sequence 93214, A
21	1752	87.3	391	US-09-724-676-93215	Sequence 93215, A
22	1752	87.3	391	US-09-724-676-93216	Sequence 93216, A
23	1752	87.3	391	US-09-724-676-93217	Sequence 93217, A
24	1752	87.3	391	US-09-724-676-93218	Sequence 93218, A
25	1752	87.3	391	US-09-724-676-93219	Sequence 93219, A
26	1752	87.3	391	US-09-724-676-93220	Sequence 93220, A

27	1752	87.3	391	US-09-724-676-93221	Sequence 93221, A
28	1752	87.3	391	US-09-724-676A-93213	Sequence 93213, A
29	1752	87.3	391	US-09-724-676A-93214	Sequence 93214, A
30	1752	87.3	391	US-09-724-676A-93215	Sequence 93215, A
31	1752	87.3	391	US-09-724-676A-93216	Sequence 93216, A
32	1752	87.3	391	US-09-724-676A-93217	Sequence 93217, A
33	1752	87.3	391	US-09-724-676A-93218	Sequence 93218, A
34	1752	87.3	391	US-09-724-676A-93219	Sequence 93219, A
35	1752	87.3	391	US-09-724-676A-93220	Sequence 93220, A
36	1752	87.3	391	US-09-724-676A-93221	Sequence 93221, A
37	1595	79.4	354	US-09-724-676-93222	Sequence 93222, A
38	1595	79.4	354	US-09-724-676-93223	Sequence 93223, A
39	1595	79.4	354	US-09-724-676-93224	Sequence 93224, A
40	1595	79.4	354	US-09-724-676-93225	Sequence 93225, A
41	1595	79.4	354	US-09-724-676-93226	Sequence 93226, A
42	1595	79.4	354	US-09-724-676-93227	Sequence 93227, A
43	1595	79.4	354	US-09-724-676-93228	Sequence 93228, A
44	1595	79.4	354	US-09-724-676-93229	Sequence 93229, A
45	1595	79.4	354	US-09-724-676-93230	Sequence 93230, A
			5	US-09-724-676-93231	Sequence 93231, A

ALIGNMENTS

RESULT 1
US-09-724-676-93212
Sequence 93212, Application US/09724676
GENERAL INFORMATION:
APPLICANT: CompuGen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 CompuGen
CURRENT APPLICATION NUMBER: US/09/724, 676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 93212
LENGTH: 430
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-93212

Query Match	99.6%	Score 2000;	DB 5;	Length 430;
Best Local Similarity	99.5%;	Pred. No. 4e-176;		
Matches 363;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	MPPAVVAESVSSLTADAFIAGESSAPTPPPALPRFFICSPDCSANSKAKLDAH	60	
DB	66	LDPPAVVAESVSSLTADAFIAGESSAPTPPPALPRFFICSPDCSANSKAKLDAH	125	
QY	61	ICKHTGERPPVCDYEGCGKAFIRDYHLSRHLITHTGEKFPVCAATGCDQKENTSNLKH	120	
DB	126	ICKHTGERPPVCDYEGCGKAFIRDYHLSRHLITHTGEKFPVCAANGCDQKENTSNLKH	185	
QY	121	FERKHENDQOKYICSPEDCKKTFKKHQOLKIHOCQHTNBPFLKCTQEGCGKHFASPSKL	180	
DB	166	FERKHENDQOKYICSPEDCKKTFKKHQOLKIHOCQHTNBPFLKCTQEGCGKHFASPSKL	245	
QY	181	RIAKKHGEYVQCKGCSVAATWTTELKHAHETKEBILCEVCRKFFKKDYLKQHMKTHA	240	
DB	246	RIAKKHGEYVQCKGCSVAATWTTELKHAHETKEBILCEVCRKFFKKDYLKQHMKTHA	305	
QY	241	PERDVCRCPRGCGCRPTTTFVFNQSHLSFHEHSRPFVCEHAGCKRTFAMQSLTRHAVV	300	
DB	306	PERDVCRCPRGCGCRPTTTFVFNQSHLSFHEHSRPFVCEHAGCKRTFAMQSLTRHAVV	365	
QY	301	HDPPKKKKKKLVKKSRSKRSLSLHSGYIPPKKQGGISLCONESPNCVEDKMLSTVA	360	
DB	366	HDPPKKKKKKLVKKSRSKRSLSLHSGYIPPKKQGGISLCONESPNCVEDKMLSTVA	425	
QY	361	VLTLG 365		
DB	426	VLTLG 430		

```
RESULT 2
US-09-724-676-93223
: Sequence 93223, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 93223
: LENGTH: 430
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-724-676-93223

Query Match
Best Local Similarity 99.6%; Score 2000; DB 5; Length 430:
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPYVAESVSLTIDAFIAGESSAPTPRPALPRFICSPFDCSANSKAMKLDH 60
DB 66 LDPYVAESVSLTIDAFIAGESSAPTPRPALPRFICSPFDCSANSKAMKLDH 125
OY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 120
DB 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 185
OY 121 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 180
DB 186 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 245
OY 181 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 240
DB 246 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 305
OY 241 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 300
DB 306 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 365
OY 301 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 360
DB 366 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
DB 426 VLTIG 430

RESULT 3
US-09-724-676-93234
: Sequence 93234, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 93234
: LENGTH: 430
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-724-676-93234

Query Match
Best Local Similarity 99.6%; Score 2000; DB 5; Length 430:
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPYVAESVSLTIDAFIAGESSAPTPRPALPRFICSPFDCSANSKAMKLDH 60
DB 66 LDPYVAESVSLTIDAFIAGESSAPTPRPALPRFICSPFDCSANSKAMKLDH 125
OY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 120
DB 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 185
OY 121 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 180
DB 186 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 245
OY 181 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 240
DB 246 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 305
OY 241 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 300
DB 306 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 365
OY 301 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 360
DB 366 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
DB 426 VLTIG 430
```

```

DB 66 LDPYVAESVSLTIDAFIAGESSAPTPRPALPRFICSPFDCSANSKAMKLDH 125
OY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 120
DB 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 185
OY 121 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 180
DB 186 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 245
OY 181 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 240
DB 246 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 305
OY 241 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 300
DB 306 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 365
OY 301 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 360
DB 366 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
DB 426 VLTIG 430

RESULT 4
US-09-724-676-93245
: Sequence 93245, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 93245
: LENGTH: 430
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-724-676-93245

Query Match
Best Local Similarity 99.6%; Score 2000; DB 5; Length 430:
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPYVAESVSLTIDAFIAGESSAPTPRPALPRFICSPFDCSANSKAMKLDH 60
DB 66 LDPYVAESVSLTIDAFIAGESSAPTPRPALPRFICSPFDCSANSKAMKLDH 125
OY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 120
DB 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHILTHGKRPVCAATGCCDOKFNTKSNLKH 185
OY 121 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 180
DB 186 FERHENDQOYICSPFDCCKTKRKHOQLIHOCQHTNEPLFKCTQOEGCGHFPSPSKL 245
OY 181 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 240
DB 246 RHANHEGYVQKGCSPVATWTLLKHVRETHKEELICEVCRKTFKRDYLRKQHMKTHA 305
OY 241 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 300
DB 306 PERDYCPCRECGGRTTYTFNLSHILSFHESRPVCEHAGCGKTFAMKOSLTRNAVY 365
OY 301 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 360
DB 366 HDPKXKMKLVKKSREKRSLSHLSGYIPRKQOGISLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
DB 426 VLTIG 430
```

OY 361 VLTIG 365
|||||
Db 426 VLTIG 430

RESULT 5
US-09-724-676-93256

; Sequence 93256, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93256
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-93256

Query Match
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPRAVAESVSLTIADAFIAAGESSAPTPPPALPRRFICSPFDCSANSKAMKLDAAH 60
:|||||
Db 66 LDPRAVAESVSLTIADAFIAAGESSAPTPPPALPRRFICSPFDCSANSKAMKLDAAH 125
:|||||
OY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKRFVCAATGCCOQKNTSNNLKH 120
:|||||
Db 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKRFVCAATGCCOQKNTSNNLKH 185
:|||||
OY 121 FERKHEHQOXYICSFEDCKKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 180
:|||||
Db 186 FERKHEHQOXYICSFEDCKKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 245
:|||||
OY 181 RHAKAHGEGYVQKGCSTFVAKTWTELLKHVRETHKEEILCEVCRKTFKKDYLKQHMKTAA 240
:|||||
Db 246 RHAKAHGEGYVQKGCSTFVAKTWTELLKHVRETHKEEILCEVCRKTFKKDYLKQHMKTAA 305
:|||||
OY 241 PERDYCRCPRGCGRTYTVFNLSHLSFHEESRPFVCEHAGCGKTFAMKOSLTRAAVY 300
:|||||
Db 306 PERDYCRCPRGCGRTYTVFNLSHLSFHEESRPFVCEHAGCGKTFAMKOSLTRAAVY 365
:|||||
OY 301 HDPQKRRKMLKVKRSRKRSLASHLSGYTPPKRQOGSLCQNGESPNCVEDKMLSTIVA 360
:|||||
Db 366 HDPQKRRKMLKVKRSRKRSLASHLSGYTPPKRQOGSLCQNGESPNCVEDKMLSTIVA 425
:|||||
OY 361 VLTIG 365
|||||
Db 426 VLTIG 430

RESULT 6
US-09-724-676-93267

; Sequence 93267, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93267
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-93267

Query Match 99.6%; Score 2000; DB 5; Length 430;

Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPRAVAESVSLTIADAFIAAGESSAPTPPPALPRRFICSPFDCSANSKAMKLDAAH 60
:|||||
Db 66 LDPRAVAESVSLTIADAFIAAGESSAPTPPPALPRRFICSPFDCSANSKAMKLDAAH 125
:|||||
OY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKRFVCAATGCCOQKNTSNNLKH 120
:|||||
Db 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKRFVCAATGCCOQKNTSNNLKH 185
:|||||
OY 121 FERKHEHQOXYICSFEDCKKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 180
:|||||
Db 186 FERKHEHQOXYICSFEDCKKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 245
:|||||
OY 181 RHAKAHGEGYVQKGCSTFVAKTWTELLKHVRETHKEEILCEVCRKTFKKDYLKQHMKTAA 240
:|||||
Db 246 RHAKAHGEGYVQKGCSTFVAKTWTELLKHVRETHKEEILCEVCRKTFKKDYLKQHMKTAA 305
:|||||
OY 241 PERDYCRCPRGCGRTYTVFNLSHLSFHEESRPFVCEHAGCGKTFAMKOSLTRAAVY 300
:|||||
Db 306 PERDYCRCPRGCGRTYTVFNLSHLSFHEESRPFVCEHAGCGKTFAMKOSLTRAAVY 365
:|||||
OY 301 HDPQKRRKMLKVKRSRKRSLASHLSGYTPPKRQOGSLCQNGESPNCVEDKMLSTIVA 360
:|||||
Db 366 HDPQKRRKMLKVKRSRKRSLASHLSGYTPPKRQOGSLCQNGESPNCVEDKMLSTIVA 425
:|||||
OY 361 VLTIG 365
|||||
Db 426 VLTIG 430

RESULT 7

US-09-724-676-93278
; Sequence 93278, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93278
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-93278

Query Match 99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPRAVAESVSLTIADAFIAAGESSAPTPPPALPRRFICSPFDCSANSKAMKLDAAH 60
:|||||
Db 66 LDPRAVAESVSLTIADAFIAAGESSAPTPPPALPRRFICSPFDCSANSKAMKLDAAH 125
:|||||
OY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKRFVCAATGCCOQKNTSNNLKH 120
:|||||
Db 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKRFVCAATGCCOQKNTSNNLKH 185
:|||||
OY 121 FERKHEHQOXYICSFEDCKKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 180
:|||||
Db 186 FERKHEHQOXYICSFEDCKKTFKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 245
:|||||
OY 181 RHAKAHGEGYVQKGCSTFVAKTWTELLKHVRETHKEEILCEVCRKTFKKDYLKQHMKTAA 240
:|||||
Db 246 RHAKAHGEGYVQKGCSTFVAKTWTELLKHVRETHKEEILCEVCRKTFKKDYLKQHMKTAA 305
:|||||
OY 241 PERDYCRCPRGCGRTYTVFNLSHLSFHEESRPFVCEHAGCGKTFAMKOSLTRAAVY 300
:|||||
Db 306 PERDYCRCPRGCGRTYTVFNLSHLSFHEESRPFVCEHAGCGKTFAMKOSLTRAAVY 365
:|||||

OY 301 HDPKMKKMLKVKKSRKRSLSHLSGYIPPRKQOGLSLCQNGESPNCVEDKMLSTVA 360
| | | | |
DB 366 HDPKMKKMLKVKKSRKRSLSHLSGYIPPRKQOGLSLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
| | | | |
DB 426 VLTIG 430

RESULT 8
US-09-724-676-93289

; Sequence 93289 Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93289
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-93289

Query Match 99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;

Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPAAVAESVSLTIADAFIAGESSAPTPPPALPRFICSPDCSANSKAMKLDH 60
| | | | |
DB 66 LDPAAVAESVSLTIADAFIAGESSAPTPPPALPRFICSPDCSANSKAMKLDH 125
OY 61 LCKHTGERPFVCDYEGCGKAFIDYHLSRHLTHGKERPFVCAANGCQKNTKSNLKH 120
| | | | |
DB 126 LCKHTGERPFVCDYEGCGKAFIDYHLSRHLTHGKERPFVCAANGCQKNTKSNLKH 185
OY 121 FERKHENOQOYICSPEDCKKTKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 180
| | | | |
DB 186 FERKHENOQOYICSPEDCKKTKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 245
OY 181 RHAAHNEGVCQKCSFVAKTWTELLKHVRETHKEEILCEVCRTFKRKDYLKQHMKTTHA 240
| | | | |
DB 246 RHAAHNEGVCQKCSFVAKTWTELLKHVRETHKEEILCEVCRTFKRKDYLKQHMKTTHA 305
OY 241 PERDYCRPREGCGRTTYTFVNLQSHILSFHEESRPVCEHAGCGKTFAMKQSLTRHAYV 300
| | | | |
DB 306 PERDYCRPREGCGRTTYTFVNLQSHILSFHEESRPVCEHAGCGKTFAMKQSLTRHAYV 365
OY 301 HDPKMKKMLKVKKSRKRSLSHLSGYIPPRKQOGLSLCQNGESPNCVEDKMLSTVA 360
| | | | |
DB 366 HDPKMKKMLKVKKSRKRSLSHLSGYIPPRKQOGLSLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
| | | | |
DB 426 VLTIG 430

RESULT 9
US-09-724-676-93300

; Sequence 93300 Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93300
; LENGTH: 430
; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-724-676-93300

Query Match 99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;

Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPAAVAESVSLTIADAFIAGESSAPTPPPALPRFICSPDCSANSKAMKLDH 60
| | | | |
DB 66 LDPAAVAESVSLTIADAFIAGESSAPTPPPALPRFICSPDCSANSKAMKLDH 125
OY 61 LCKHTGERPFVCDYEGCGKAFIDYHLSRHLTHGKERPFVCAANGCQKNTKSNLKH 120
| | | | |
DB 126 LCKHTGERPFVCDYEGCGKAFIDYHLSRHLTHGKERPFVCAANGCQKNTKSNLKH 185
OY 121 FERKHENOQOYICSPEDCKKTKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 180
| | | | |
DB 186 FERKHENOQOYICSPEDCKKTKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 245
OY 181 RHAAHNEGVCQKCSFVAKTWTELLKHVRETHKEEILCEVCRTFKRKDYLKQHMKTTHA 240
| | | | |
DB 246 RHAAHNEGVCQKCSFVAKTWTELLKHVRETHKEEILCEVCRTFKRKDYLKQHMKTTHA 305
OY 241 PERDYCRPREGCGRTTYTFVNLQSHILSFHEESRPVCEHAGCGKTFAMKQSLTRHAYV 300
| | | | |
DB 306 PERDYCRPREGCGRTTYTFVNLQSHILSFHEESRPVCEHAGCGKTFAMKQSLTRHAYV 365
OY 301 HDPKMKKMLKVKKSRKRSLSHLSGYIPPRKQOGLSLCQNGESPNCVEDKMLSTVA 360
| | | | |
DB 366 HDPKMKKMLKVKKSRKRSLSHLSGYIPPRKQOGLSLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
| | | | |
DB 426 VLTIG 430

RESULT 10

US-09-724-676A-93212
; Sequence 93212 Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93212
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-93212

Query Match 99.6%; Score 2000; DB 5; Length 430;
Best Local Similarity 99.5%; Pred. No. 4e-176;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDPAAVAESVSLTIADAFIAGESSAPTPPPALPRFICSPDCSANSKAMKLDH 60
| | | | |
DB 66 LDPAAVAESVSLTIADAFIAGESSAPTPPPALPRFICSPDCSANSKAMKLDH 125
OY 61 LCKHTGERPFVCDYEGCGKAFIDYHLSRHLTHGKERPFVCAANGCQKNTKSNLKH 120
| | | | |
DB 126 LCKHTGERPFVCDYEGCGKAFIDYHLSRHLTHGKERPFVCAANGCQKNTKSNLKH 185
OY 121 FERKHENOQOYICSPEDCKKTKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 180
| | | | |
DB 186 FERKHENOQOYICSPEDCKKTKKHQOLKHOCQHTNEPLFKCTQEGCGKHFASPSKLK 245
OY 181 RHAAHNEGVCQKCSFVAKTWTELLKHVRETHKEEILCEVCRTFKRKDYLKQHMKTTHA 240
| | | | |
DB 246 RHAAHNEGVCQKCSFVAKTWTELLKHVRETHKEEILCEVCRTFKRKDYLKQHMKTTHA 305


```
QY 241 PERDVCRCPRGCGRTTYVFNLOSHILSFHEESRPVCEHAGCKTAMKQSLTRHAYV 300
; SEQUENCE: 93223
; SEQ ID NO 93234
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93223
Db 306 PERDVCRCPRGCGRTTYVFNLOSHILSFHEESRPVCEHAGCKTAMKQSLTRHAYV 365
QY 301 HDPPKAKKMLKVKRSRKSRLASHLSGYIPPKRKOQGLSLCONEESPNCVEDKMLSTVA 360
; SEQUENCE: 93223
; SEQ ID NO 93234
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93223
Db 366 HDPPKAKKMLKVKRSRKSRLASHLSGYIPPKRKOQGLSLCONEESPNCVEDKMLSTVA 425
QY 361 VLTUG 365
; SEQUENCE: 93223
; SEQ ID NO 93234
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93223
Db 426 VLTUG 430
```

```
RESULT 11
US-09-724-676A-93223
; Sequence 93223, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93223
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93223
```

```
Query Match
Best Local Similarity 99.5%; Score 2000; DB 5; Length 430;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPDCSANSKAMKLDAAH 60
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 66 LDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPDCSANSKAMKLDAAH 125
QY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDOKNTKSNLKKH 120
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDOKNTKSNLKKH 185
QY 121 FERKHEHNOQKQYICSPFEDCKKTKKHQOLKHOCQHTNPLFKCTQEGCGHFPASPSKX 180
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 186 FERKHEHNOQKQYICSPFEDCKKTKKHQOLKHOCQHTNPLFKCTQEGCGHFPASPSKX 245
QY 181 RHAAHNEGYVQKGCSPVATWTETLLKHVRETHKEEILCEVCRKTFKRDYLRKQHMKTHA 240
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 246 RHAAHNEGYVQKGCSPVATWTETLLKHVRETHKEEILCEVCRKTFKRDYLRKQHMKTHA 305
QY 241 PERDVCRCPRGCGRTTYVFNLOSHILSFHEESRPVCEHAGCKTAMKQSLTRHAYV 300
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 306 PERDVCRCPRGCGRTTYVFNLOSHILSFHEESRPVCEHAGCKTAMKQSLTRHAYV 365
QY 301 HDPPKAKKMLKVKRSRKSRLASHLSGYIPPKRKOQGLSLCONEESPNCVEDKMLSTVA 360
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 366 HDPPKAKKMLKVKRSRKSRLASHLSGYIPPKRKOQGLSLCONEESPNCVEDKMLSTVA 425
QY 361 VLTUG 365
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 426 VLTUG 430
```

```
RESULT 12
US-09-724-676A-93234
; Sequence 93234, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93234
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93234
```

```
Query Match
Best Local Similarity 99.5%; Score 2000; DB 5; Length 430;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPDCSANSKAMKLDAAH 60
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 66 LDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPDCSANSKAMKLDAAH 125
QY 61 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDOKNTKSNLKKH 120
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 126 LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLITHTGEKPFVCAATGCDOKNTKSNLKKH 185
QY 121 FERKHEHNOQKQYICSPFEDCKKTKKHQOLKHOCQHTNPLFKCTQEGCGHFPASPSKX 180
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 186 FERKHEHNOQKQYICSPFEDCKKTKKHQOLKHOCQHTNPLFKCTQEGCGHFPASPSKX 245
QY 181 RHAAHNEGYVQKGCSPVATWTETLLKHVRETHKEEILCEVCRKTFKRDYLRKQHMKTHA 240
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 246 RHAAHNEGYVQKGCSPVATWTETLLKHVRETHKEEILCEVCRKTFKRDYLRKQHMKTHA 305
QY 241 PERDVCRCPRGCGRTTYVFNLOSHILSFHEESRPVCEHAGCKTAMKQSLTRHAYV 300
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 306 PERDVCRCPRGCGRTTYVFNLOSHILSFHEESRPVCEHAGCKTAMKQSLTRHAYV 365
QY 301 HDPPKAKKMLKVKRSRKSRLASHLSGYIPPKRKOQGLSLCONEESPNCVEDKMLSTVA 360
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 366 HDPPKAKKMLKVKRSRKSRLASHLSGYIPPKRKOQGLSLCONEESPNCVEDKMLSTVA 425
QY 361 VLTUG 365
; SEQUENCE: 93245
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Db 426 VLTUG 430
```

```
RESULT 13
US-09-724-676A-93245
; Sequence 93245, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 93245
; LENGTH: 430
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-724-676A-93245
Query Match
Best Local Similarity 99.5%; Score 2000; DB 5; Length 430;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 181 RHAHAHEGYVOCSCFVAKTWELKLVHRETHKEEILCEVCRKTFKRRDYLKOHMKTNA 240
      |||||||
Db 246 RHAHAHEGYVOCSCFVAKTWELKLVHRETHKEEILCEVCRKTFKRRDYLKOHMKTNA 305
OY 241 PERDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAVV 300
      |||||||
Db 306 PERDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAVV 365
OY 301 HDPDKKKMKLKVYKKSREKRSLSHLSGYIPPKRKOQGLSLCQNGESPNCVEDKMLSTVA 360
      |||||||
Db 366 HDPDKKKMKLKVYKKSREKRSLSHLSGYIPPKRKOQGLSLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
      |||||
Db 426 VLTIG 430
```

RESULT 14

```
US-09-724-676A-93256
; Sequence 93256, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ. ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ. ID NO 93256
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-93256
```

Query Match 99.6%; Score 2000; DB 5; Length 430;

Best Local Similarity 99.5%; Pred. No. 4e-176; Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
OY 1 MDPPAVVAESVSLITIDAFIAGESSAPTPRRPALPRFFICSPDCCSANTSKAMKIDAH 60
      |||||||
Db 66 LDPPAVVAESVSLITIDAFIAGESSAPTPRRPALPRFFICSPDCCSANTSKAMKIDAH 125
OY 61 LCKHTGERPEVCDYEGCGKAFIRDYHLSRHILTHGEEKPFVCAANGCDQKFNKSNLKKH 120
      |||||||
Db 126 LCKHTGERPEVCDYEGCGKAFIRDYHLSRHILTHGEEKPFVCAANGCDQKFNKSNLKKH 185
OY 121 FERKHEHNOQKOYICSPEDCKKTFKKHOOLKIHQCOHTNRPFLKCTOEGCGKHFPASPSKLK 180
      |||||||
Db 186 FERKHEHNOQKOYICSPEDCKKTFKKHOOLKIHQCOHTNRPFLKCTOEGCGKHFPASPSKLK 245
OY 181 RHAHAHEGYVOCSCFVAKTWELKLVHRETHKEEILCEVCRKTFKRRDYLKOHMKTNA 240
      |||||||
Db 246 RHAHAHEGYVOCSCFVAKTWELKLVHRETHKEEILCEVCRKTFKRRDYLKOHMKTNA 305
OY 241 PERDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAVV 300
      |||||||
Db 306 PERDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAVV 365
OY 301 HDPDKKKMKLKVYKKSREKRSLSHLSGYIPPKRKOQGLSLCQNGESPNCVEDKMLSTVA 360
      |||||||
Db 366 HDPDKKKMKLKVYKKSREKRSLSHLSGYIPPKRKOQGLSLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
      |||||
Db 426 VLTIG 430
```

RESULT 15

```
US-09-724-676A-93267
; Sequence 93267, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
```

```
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ. ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ. ID NO 93267
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-93267
```

Query Match 99.6%; Score 2000; DB 5; Length 430;

Best Local Similarity 99.5%; Pred. No. 4e-176; Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
OY 1 MDPPAVVAESVSLITIDAFIAGESSAPTPRRPALPRFFICSPDCCSANTSKAMKIDAH 60
      |||||||
Db 66 LDPPAVVAESVSLITIDAFIAGESSAPTPRRPALPRFFICSPDCCSANTSKAMKIDAH 125
OY 61 LCKHTGERPEVCDYEGCGKAFIRDYHLSRHILTHGEEKPFVCAANGCDQKFNKSNLKKH 120
      |||||||
Db 126 LCKHTGERPEVCDYEGCGKAFIRDYHLSRHILTHGEEKPFVCAANGCDQKFNKSNLKKH 185
OY 121 FERKHEHNOQKOYICSPEDCKKTFKKHOOLKIHQCOHTNRPFLKCTOEGCGKHFPASPSKLK 180
      |||||||
Db 186 FERKHEHNOQKOYICSPEDCKKTFKKHOOLKIHQCOHTNRPFLKCTOEGCGKHFPASPSKLK 245
OY 181 RHAHAHEGYVOCSCFVAKTWELKLVHRETHKEEILCEVCRKTFKRRDYLKOHMKTNA 240
      |||||||
Db 246 RHAHAHEGYVOCSCFVAKTWELKLVHRETHKEEILCEVCRKTFKRRDYLKOHMKTNA 305
OY 241 PERDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAVV 300
      |||||||
Db 306 PERDVCRCPREGCGRTYTVFNLSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAVV 365
OY 301 HDPDKKKMKLKVYKKSREKRSLSHLSGYIPPKRKOQGLSLCQNGESPNCVEDKMLSTVA 360
      |||||||
Db 366 HDPDKKKMKLKVYKKSREKRSLSHLSGYIPPKRKOQGLSLCQNGESPNCVEDKMLSTVA 425
OY 361 VLTIG 365
      |||||
Db 426 VLTIG 430
```

Search completed: February 10, 2003, 17:50:24

Job time : 42 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 12:58:51 : Search time 70 Seconds
(without alignments)
694.807 Million cell updates/sec

Title: US-09-831-426c-2

Perfect score: 2008
Sequence: 1 MDPPEVAEVSYSLTITDAF.....ESPNCVEDKMLSTVAVLTIIG 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq..101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2008	100.0	365	21	AAV93317
2	2000	99.6	409	22	AAV93317
3	1933	96.3	423	17	AAV93317
4	821	40.9	184	23	ABP41846
5	563.5	28.1	220	22	AAV93317
6	453	22.6	95	22	ABG03947
7	425	21.0	927	22	AAV93317
8	421	21.0	869	22	AAV93317
9	417.5	20.8	803	22	AAV93317
10	414.5	20.6	904	22	AAV93317

11	410.5	20.4	632	22	AAV93317	Human protein SEQ
12	410.5	20.4	631	22	AAV93317	Human protein SEQ
13	410.5	20.4	632	22	ABG16386	Novel human diagno
14	410.5	20.4	652	22	AAV93317	Human protein SEQ
15	406.5	20.2	403	22	ABG64531	Drosophila melanog
16	406.5	20.2	700	22	AAV93317	Human polypeptide
17	405	20.2	1241	22	ABG02797	Novel human diagno
18	404	20.1	779	22	ABG00880	Novel human diagno
19	404	20.1	779	22	AAV93317	Human protein SEQ
20	404	20.1	779	22	AAV93317	Human protein SEQ
21	403	20.1	934	22	ABG69958	Drosophila melanog
22	398.5	19.8	773	22	AAV93317	Human polypeptide
23	396.5	19.7	646	22	ABP51377	Human MDR1 seq ID
24	396	19.7	577	22	AAV93317	Human polypeptide
25	395.5	19.7	412	21	AAV93317	A transcription fact
26	395.5	19.7	1050	22	ABG09685	Novel human diagno
27	395	19.7	803	22	ABG95278	Human protein sequ
28	394.5	19.6	727	22	ABG16954	Novel human diagno
29	394.5	19.6	809	22	AAV93317	Human polypeptide
30	393.5	19.6	751	22	AAV93317	Human polypeptide
31	393	19.6	489	22	ABG50238	Human transcription
32	393	19.6	702	23	ABG04333	Human nucleic acid
33	393	19.6	799	21	ABG21003	Human protein SEQ
34	392.5	19.5	547	22	AAV93317	Human protein SEQ
35	392.5	19.5	570	22	AAV93317	Human protein SEQ
36	391.5	19.5	632	22	ABG19011	Novel human diagno
37	388.5	19.3	817	22	AAV93317	Human polypeptide
38	387.5	19.3	1252	23	AAV93317	Human protein SEQ
39	386.5	19.2	759	23	AAV93317	Mouse SCAN/KRAB pr
40	386	19.2	406	22	AAV93317	Human novel secret
41	385.5	19.2	577	23	AAV93317	Human transcription
42	385.5	19.2	1520	22	ABG27130	Novel human diagno
43	384.5	19.1	357	22	AAV93317	Human protein sequ
44	384.5	19.1	366	22	AAV93317	Human polypeptide
45	384.5	19.1	613	22	AAV93317	Human protein sequ

ALIGNMENTS

RESULT 1
AAV93317
ID AAV93317 standard; Protein; 365 AA.
XX
AC AAV93317;
XX
DT 04-SEP-2000 (first entry)
XX
DE A human transcription factor designated htfiia.
XX
KW Human; transcription factor; htfiia; DNA-binding protein;
KW transcription; ribosomal RNA 5S gene; transcriptional control;
KW cancer.
XX
OS Homo sapiens.
XX
PN WO200028024-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR02738.
XX
PR 10-NOV-1998; 98FR-0014146.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Rocher C;
XX
DR WPI: 2000-387419/33.
XX
PT N-PSDB; AAA15405.
XX
PT New nucleic acid encoding human transcription factor htfiia, useful for treatment and diagnosis of cancer and inherited disease

XX Claim 10; Page 40-41; 49pp; French.

PS The present sequence represents a human transcription factor designated

XX htfi1a. The polypeptide is probably a DNA-binding protein probably

CC involved in initiating transcription of the gene for ribosomal RNA 5S

CC and maintaining the stability of transcription of other control genes.

CC The htfi1a polynucleotides and polypeptides are used to make

CC therapeutic or diagnostic compositions for diseases associated with

CC disorders of transcriptional control, particularly cancer or other

CC inherited diseases. The htfi1a polynucleotide can also be used to

CC detect anomalies in gene transcription, particularly for diagnosis of

CC inherited disease, also for studying diseases involving htfi1a.

XX

XX Sequence 365 AA:

Query Match 100.0%; Score 2008; DB 21; Length 365;

Best Local Similarity 100.0%; Pred. No. 2.5e-159;

Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDPAAVVAESVSLTIADAFIAGESSAPTPRPALPRRFICSPFDCSANYSKAMKLDAN 60

Db 1 MDPAAVVAESVSLTIADAFIAGESSAPTPRPALPRRFICSPFDCSANYSKAMKLDAN 60

QY 61 LCKHTGERPVCDEGCGKAFIDYHLSRHILTHGERPVCVAATGCDOKNTKSNLKH 120

Db 61 LCKHTGERPVCDEGCGKAFIDYHLSRHILTHGERPVCVAATGCDOKNTKSNLKH 120

QY 121 FERKHEMOOKOYICSEFEDCKKTFKKHQOLKHQOCHNEPRLFKCTOEGCGNHFASPSKLK 180

Db 121 FERKHEMOOKOYICSEFEDCKKTFKKHQOLKHQOCHNEPRLFKCTOEGCGNHFASPSKLK 180

QY 181 RHAKAHGEGYCOQKCSFVATWTLLKHYRETHKEEILCEVCGRKFKKKDYLKQHMKTNA 240

Db 181 RHAKAHGEGYCOQKCSFVATWTLLKHYRETHKEEILCEVCGRKFKKKDYLKQHMKTNA 240

QY 241 PERDYCRCPREGCGRTTYTFVNLQSHILSFHEESRPFCVCEHAGCGKTFAMKQSLTRNAVY 300

Db 241 PERDYCRCPREGCGRTTYTFVNLQSHILSFHEESRPFCVCEHAGCGKTFAMKQSLTRNAVY 300

QY 301 HDPPKKMKLKVKKSRKRSLSHLSGYTPRKQOGGLSCQNGESPNCVEDMLSTVA 360

Db 301 HDPPKKMKLKVKKSRKRSLSHLSGYTPRKQOGGLSCQNGESPNCVEDMLSTVA 360

QY 361 VLTIG 365

Db 361 VLTIG 365

RESULT 2

AAG75181

ID AAG75181 standard; Protein; 409 AA.

XX AAG75181;

AC

DT 03-SEP-2001 (first entry)

XX

DE Human colon cancer antigen protein SEQ ID NO:5945.

XX

KM Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; Chromosome 13.

XX

OS Homo sapiens.

XX

PN MO200122920-A2.

XX

PD 05-APR-2001.

XX

PP 28-SEP-2000; 2000MD-US26524.

XX

PR 29-SEP-1999; 99US-0157137.

XX

PR 03-NOV-1999; 99US-0163280.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

DR WPI: 2001-235357/24.

XX

DR N-PSDB: AAH34586.

XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX

PS Claim 11; Page 7452-7454; 9803pp; English.

XX

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where

CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders

CC associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patient's own production of P.

CC Additionally, N may be used to produce the colon cancer-associated P,

CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC and AAB77789 represent sequences used in the exemplification of the

CC present invention.

CC

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX

XX sequence 409 AA:

Query Match 99.6%; Score 2000; DB 22; Length 409;

Best Local Similarity 99.5%; Pred. No. 1.4e-158;

Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPAAVVAESVSLTIADAFIAGESSAPTPRPALPRRFICSPFDCSANYSKAMKLDAN 60

Db 45 LDPAAVVAESVSLTIADAFIAGESSAPTPRPALPRRFICSPFDCSANYSKAMKLDAN 104

QY 61 LCKHTGERPVCDEGCGKAFIDYHLSRHILTHGERPVCVAATGCDOKNTKSNLKH 120

Db 61 LCKHTGERPVCDEGCGKAFIDYHLSRHILTHGERPVCVAATGCDOKNTKSNLKH 120

QY 105 LCKHTGERPVCDEGCGKAFIDYHLSRHILTHGERPVCVAATGCDOKNTKSNLKH 164

Db 105 LCKHTGERPVCDEGCGKAFIDYHLSRHILTHGERPVCVAATGCDOKNTKSNLKH 164

QY 121 FERKHEMOOKOYICSEFEDCKKTFKKHQOLKHQOCHNEPRLFKCTOEGCGNHFASPSKLK 180

Db 121 FERKHEMOOKOYICSEFEDCKKTFKKHQOLKHQOCHNEPRLFKCTOEGCGNHFASPSKLK 180

QY 165 FERKHEMOOKOYICSEFEDCKKTFKKHQOLKHQOCHNEPRLFKCTOEGCGNHFASPSKLK 224

Db 165 FERKHEMOOKOYICSEFEDCKKTFKKHQOLKHQOCHNEPRLFKCTOEGCGNHFASPSKLK 224

QY 181 RHAKAHGEGYCOQKCSFVATWTLLKHYRETHKEEILCEVCGRKFKKKDYLKQHMKTNA 240

Db 225 RHAKAHGEGYCOQKCSFVATWTLLKHYRETHKEEILCEVCGRKFKKKDYLKQHMKTNA 284

QY 241 PERDYCRCPREGCGRTTYTFVNLQSHILSFHEESRPFCVCEHAGCGKTFAMKQSLTRNAVY 300

Db 285 PERDYCRCPREGCGRTTYTFVNLQSHILSFHEESRPFCVCEHAGCGKTFAMKQSLTRNAVY 344

QY 301 HDPPKKMKLKVKKSRKRSLSHLSGYTPRKQOGGLSCQNGESPNCVEDMLSTVA 360

Db 345 HDPPKKMKLKVKKSRKRSLSHLSGYTPRKQOGGLSCQNGESPNCVEDMLSTVA 404

QY 361 VLTIG 365

Db 405 VLTIG 409

RESULT 3

AAR91305

ID AAR91305 standard; Protein; 423 AA.

XX AAR91305;

AC

AC AAR91305;

XX

DT 07-JUL-1996 (first entry)
 XX
 DE Transcription factor-IIIA.
 XX
 KW Human; transcription factor-IIIA; hTFIIIA; DNA binding protein;
 KW ribosome; zinc finger; diagnostic; probe; transcription control;
 KW antitumour; cancer; therapy.
 XX
 OS Homo sapiens.
 XX
 PM EP704526-A1.
 PD 03-APR-1996.
 XX
 PF 05-SEP-1995; 95EP-0113908.
 XX
 PR 05-SEP-1994; 94JP-0211022.
 XX
 PA (SAKA) OTSUKA PHARM CO LTD.
 XX
 PI Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;
 DR WPI: 1996-173033/18.
 DR N-PSDB: AAT14037, AAT14038.
 PT Human Transcription Factor III A gene - useful in regulation of
 PT transcription and for diagnosis and treatment of e.g. cancer related
 PT diseases
 XX
 PS Claim 1; Page 8-10; 17pp; English.
 XX
 CC The sequence represents human transcription factor-IIIA (hTFIIIA).
 CC a DNA binding protein which is necessary for the initiation of 5S
 CC RNA gene transcription, binding to an internal control region of
 CC the 3S gene. The protein contains 9 zinc finger domains, which are
 CC homologous to the C2H2 finger domains of Xenopus TFIIIA, except for
 CC the 6th finger domain, which has only 3 amino acid residues between
 CC 2 cysteine residues, instead of 5 amino acid residues for Xenopus
 CC TFIIIA. The protein (optionally in recombinant form) and encoding
 CC gene may be used in diagnosis, identification or therapy of
 CC hereditary diseases such as cancer, or other diseases resulting from
 CC abnormal transcriptional control, and to analyse the mechanisms
 CC involved in their activity.
 CC
 XX
 SQ Sequence 423 AA:
 Query Match 96.3%; Score 1933; DB 17; Length 423;
 Best Local Similarity 96.2%; Pred. No. 5,5e-153;
 Matches 351; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 0;
 QY 1 MDPRAVVAESVSSITIDAFIAGESSAPTPPPALPRPFTCSFDCSANTSKAMKLDAN 60
 :|||||
 DB 59 LDPRAVVAESVSSITIDAFIAGESSAPTPPPALPRPFTCSFDCSANTSKAMKLDAN 118
 :|||||
 QY 61 LCKHTGERPFCVDEGGCGKAFIRDYHLSRHILTHTEGKFFVCAACGQCKNTSKNKKH 120
 :|||||
 DB 119 LCKHTGERPFCVDEGGCGKAFIRDYHLSRHILTHTEGKFFVCAACGQCKNTSKNKKH 178
 :|||||
 QY 121 FERKHEHQNOYICSFDCCKTKTKKQOLKIIOCHTNPRLEKCTQEGCGKHFASPSKTK 180
 :|||||
 DB 179 FERKHEHQNOYICSFDCCKTKTKKQOLKIIOCHTNPRLEKCTQEGCGKHFASPSKTK 238
 :|||||
 QY 181 RHANAHEGYVQCKCSFVAATWTLELKHVRETHKEEILCEVCRKTFKKRDYLOHMKTHA 240
 :|||||
 DB 239 RHANAHEGYVQCKCSFVAATWTLELKHVRETHKEEILCEVCRKTFKKRDYLOHMKTHA 298
 :|||||
 QY 241 PERDYVCRPREGGCGRTTTFVFNOSHIISFHESRFRVCEHAGCKRTKAMQSLTRNAV 300
 :|||||
 DB 299 PERDYVCRPREGGCGRTTTFVFNOSHIISFHESRFRVCEHAGCKRTKAMQSLTRNAV 358
 :|||||
 QY 301 HDPKPKKKMLKVKKSRKRSILASHLSGYIPPKKQOGSLCQNESEPNCEVDMKLSIVA 360
 :|||||
 DB 359 HDPKPKKKMLKVKKSRKRSILASHLSGYIPPKKQOGSLCQNESEPNCEVDMKLSIVA 418
 :|||||

QY 361 VLTIG 365
 :|||||
 DB 419 VLTIG 423
 :|||||
 RESULT 4
 ID ABP41846 standard; Protein; 184 AA.
 ABP41846
 XX
 AC ABP41846;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HUSYA18, SEQ ID NO:2978.
 XX
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive; chromosome 13q12.3-13.1.
 XX
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-0518569.
 XX
 PR 07-JUN-2000; 2000US-209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birst CE, Rosen CA;
 XX
 DR WPI: 2002-147878/19.
 DR N-PSDB: ABQ54923.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11; SEQ ID NO 2978; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides; antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies), autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the

PT biodiversity -
XX
PS Claim 20: SEQ ID NO 39706; 103pp; English.
CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AB60010-AB630377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIP0
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SO Sequence 95 AA;
XX
Query Match 22.6%; Score 453; DB 22; Length 95;
Best Local Similarity 91.6%; Pred. No. 2.4e-30;
Matches 87; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 271 HEERPVCEHAGCGKTFAMKOSLTRHAVHPDPKKMKLYKRSREKRSLSHLSGYIP 330
Db 1 HEERPVCEHAGCGKTFAMKPSLTSDAYVHPGKKMKLYKRSREKRSLSHLSGYIP 60
QY 331 PKRQGGSLCQNGESPNCVEDKMLSTVAVLTLG 365
Db 61 PKRQGGSLCQNGESPNCVEDKMLSAVPLTLG 95
XX
RESULT 7
AAM80283
ID AAM80283 standard: Protein; 927 AA.
XX
AC AAM80283;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3929.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; Leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN MO200157190-A2.
XX
PD 09-AUG-2001.
XX
PE 05-FEB-2001; 2001MO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX

PA (HYSE-) HXSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
DR WPI: 2001-476283/51.
DR N-PSDB: AAK53416.
XX
PT Nucleic acid encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20: Page 470-471; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SO Sequence 927 AA;
XX
Query Match 21.2%; Score 425; DB 22; Length 927;
Best Local Similarity 32.0%; Pred. No. 9.1e-27;
Matches 116; Conservative 42; Mismatches 115; Indels 89; Gaps 19;
QY 3 PPAVVAESVSLTTADAFIAAGSSAFTPPRA-LPRRFICSPDCCSANYSKAMKIDAH 61
Db 183 PPVPAVNAEPLRELVQG-----RPAAGAKPYICN--ECGRSFSQWSKYLHQ 227
QY 62 CKHNGERPFVCDYEGCGKAFIRDYHLSRHILFTGKRPVC----- 102
Db 228 RHNGERPNYC--SECKSFTQSSHLYQHRTHTGKRPKCDCKGCSWSNLYQHRT 285
QY 103 -----AATGCDQKFNFKNLKKHFERKHENOQOYICSPEDCKRTKKHQOQKIHQ 155
Db 286 HTGKRPYKCTCEKAFQSTNLTKH-QRSHTG-EKPYKCG--ECRRAFYRSSDLQHQA 341
QY 156 HTNEPLFCPTQEGGKHFAPSKLRRAKHAEG--YVQK-GGSPV-----AKWT 203
Db 342 HTGKRPYKCP--CGKRFQGNHNLKHOKIHAGEKPYRTCEGKCFIOSSELTQHQRHT 399
QY 204 -----ELIKHVRETHKEE--ILCEVCRKTPKRDYLOKMTNAPER 243
Db 400 GEKPYECLECKSGFSGHSSTLIKQR-THLEDFKCPVCGKFTLSATLLRHQRHTGER 458
QY 244 DYCRCPRREGGRTYTFVFNQSHLSFHEERSPFVEHNGCGKTFAMQSLTRHAYVNDP 303
Db 459 PY-KCP--CGKSFVSNNLINH-QRIHGERPYIC--ADCKSFTMSSTLIRQRHTG 512
QY 304 DK 305
Db 513 EK 514
XX
RESULT 8
AAM79299
ID AAM79299 standard: Protein; 869 AA.
XX
AC AAM79299;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1961.

XX human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-APR-2001.
XX
PF 05-FEB-2001; 2001MO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0665361.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA {HYSE-} HYSEQ INC.
PI Zhang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
PI Tang QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejberman T, Goodrich R;
XX
WP1: 2001-476283/51.
DR N-PDSB: AAK52432.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 4372-4374; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

XX
SO Sequence 869 AA:

Query Match 21.0%; Score 421; DB 22; Length 869;
Best Local Similarity 32.0%; Ped. No. 1.8e-26;
Matches 116; Conservative 41; Mismatches 116; Indels 89; Gaps 19;

OY 3 PPAVAVESVSLIADAFIAAGSSAAPTPPPA-LPRLFISPFDCSANYSKAWKDANIL 61
DB PRPVAPNEPSRLRELVOG-----PRAGEKRYLON--ECKSFSQMSKLRIHQ 169
OY 62 CKRTGEPRPCDYEGCGKAFFIRDYLSRLITLTGCEKPPVC----- 102
DB RIHTGERPNTC--SEGCKSFSTQSHLYOHORHNGEKKYKCPDCKGCFSSWSNLYQHQT 227
OY 103 -----AATGCDOKFNFTSNLKIHFERKENENODKOYICSPEDCKTFPKKHQDKIHOC 155
DB HTGEERYKCETECIELATQSYINLIK-HQSHTG-EKPYKSG--ECRRAFYRSSDLIQHQT 283
OY 156 HTNPRLFKCTQEGCGGNFASPSKLRKNHAKNEG---TVGCK-GCSFY-----AKWT 203
DB 284 HTEGRKPYKSE--CGRRFGONHNLLKHOIKTAGEKRPYCSTGSGSFLQSSLETOHQRTH 341

OY	204	-----ELKHVRETHKEE--ILCEVCRTKFRKDYLKONHKHTAPER	243
Dd	342	GEPPTCLECGSGSPHSSTLIKNOR-THLRDEPFPCPCQGTFLSLATLRLHRRITIGER	400
OY	244	DVCRPREGCGTYYTFENLOSILSFHEBSRPFCCEIAGCGKTFAMKOSLTJRHAVNDP	303
Dd	401	PV-KCPE-CGKSFSVSNNLNH-VRIHQERPYC-ADCGKSFIMSSTLIRHORIHNG	454
OY	304	DK 305	:
Dd	455	EK 456	:
 RESULT 9 AAM79014 standard: Protein; 803 AA. ID AAM79014 AAAM79014: AAAM79014: 06-NOV-2001 (first entry) Human protein SEQ ID NO 1676. XX DT Human PX Human cytokine; cell proliferation; cell differentiation; gene therapy; KW vaccinia; peptide therapy; stem cell growth factor; haematopoiesis; KW tissue growth factor; immunomodulatory; cancer; leukaemia; KW nervous system disorder; arthritis; inflammation. XX OS Homo sapiens. XX PN MO200157190-AZ. XX PD 09-AUG-2001. XX XX XX 05-FEB-2001; 2001WO-USO4098. XX PR 03-FEB-2000; 2000US-0456914. PR 27-APR-2000; 2000US-0560875. PR 20-JUN-2000; 2000US-0598075. PR 19-JUL-2000; 2000US-0620325. PR 01-SEP-2000; 2000US-0654936. PR 15-SEP-2000; 2000US-0653861. PR 20-OCT-2000; 2000US-0653325. PR 30-NOV-2000; 2000US-0728422. XX PA (HVSE-) HVSEQ INC. XX Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y; PI Zhao QX, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW; PI Xue AJ, Yang Y, Wejhtman T, Goodrich R; DR WPI: 2001-476283/51. XX N-PSDB: AAK52147. XX PT Nucleic acids encoding polypeptides with cytokine-like activities, XX useful in diagnosis and gene therapy - XX PS Claim 20: Page 4019-4021; 6221pp; English. XX XX The invention relates to polynucleotides (AAK51456-AAK53435) and the CC encoded polypeptides (AAM78333-AAAB0302) that exhibit activity elating to CC cytokine, cell proliferation or cell differentiation or which may induce CC production of other cytokines in other cell populations. The CC polynucleotides and polypeptides are useful in gene therapy, vaccines or CC peptide therapy. The polypeptides have various cytokine-like activities, CC e.g. stem cell growth factor activity, haematopoiesis regulating CC activity, tissue growth factor activity, immunomodulatory activity and CC activin/inhibin activity and may be useful in the diagnosis and/or CC treatment of cancer, leukaemia, nervous system disorders, arthritis and CC inflammation. CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 CC (AAMB00200) are omitted as the relevant pages from the sequence listing			

KW nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Zhang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Tang QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 DR WPI: 2001-476283/51.
 DR N-PSDB: AAK52080.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 XX useful in diagnosis and gene therapy -
 XX
 PS Claim 20; Page 3945-3946; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, hematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC
 CC Note: Records for SEQ ID NO.2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM980020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SO Sequence 622 AA:

[illegible]

OY	307	KMRKLVKSKSEKSLASHLSGYIPPRKGGQGLSLCON-----GESPNCVDDKMLST	358
		: : : : : : : : : : : :	
Db	484	KCEGCGAFNMOSSSLRKH-----KRIRHGEKPYGCEGCGAFNMOSSLIRKRIHT	534
		: : : : : : : : : : : :	
RESULT 12			
XX	ID	AAW79931 standard; Protein: 631 AA.	
XX	AC	AAW79931	
XX	AA	AAW79931;	
XX	DT	06-NOV-2001 (first entry)	
XX	DE	Human protein SEQ ID NO 3577.	
XX	DE		
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;		
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;		
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;		
KW	nervous system disorder; arthritis; inflammation.		
XX			
OS	Homo sapiens.		
XX			
PN	MO200157190-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	05-FEB-2001; 2001MO-US04098.		
XX			
PR	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
PR	20-JUN-2000; 2000US-0598075.		
PR	19-JUL-2000; 2000US-0620325.		
PR	01-SEP-2000; 2000US-0654936.		
PR	15-SEP-2000; 2000US-0663561.		
PR	20-OCT-2000; 2000US-0693325.		
PR	30-NOV-2000; 2000US-0728422.		
XX			
PA	(HYSE-) HYSEO INC.		
XX			
PL	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;		
PL	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;		
PL	Xue AJ, Yang Y, Wejhrman T, Goodrich R;		
XX			
DR	WPI: 2001-476283/51.		
DR	NP-PSDB; AAK53064.		
XX			
PT	Nucleic acids encoding polypeptides with cytokine-like activities,		
XX	useful in diagnosis and gene therapy -		
XX	Claim 20; Page 388; 6221pp; English.		

XX	The invention relates to polynucleotides (AAK51456-AAK53435) and the				
CC	encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to				
CC	cytokine, cell proliferation or cell differentiation or which may induce				
CC	production of other cytokines in other cell populations. The				
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or				
CC	peptide therapy. The polypeptides have various cytokine-like activities,				
CC	e.g. stem cell growth factor activity, haematopoiesis regulating				
CC	activity, tissue growth factor activity, immunomodulatory activity and				
CC	activin/inhibin activity and may be useful in the diagnosis and/or				
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and				
CC	inflammation.				
CC	Note: Records for SFO ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666				
CC	(AAW80020) are omitted as the relevant pages from the sequence listing				
CC	were missing at the time of publication.				
XX					
XX	Sequence 631 AA:				
	Query Match	20.4%	Score 410.5	DB 22	Length 631
	Best Local Similarity	30.5%	Pred. NO. 9.1e-26		
	Matches 109; Conservative	48;	Mismatches 137;	Indels 63;	Gaps 17

QY 40 FICSPDCCSANYSKAMKLDALHCKHTGERPFVCDYEGCGKAFIRDYHLSRIHLTHTEKPY 99
 Db 212 YICE--ECGKAFYSSALNTHKRIHTGEKPYKCD--KCDKAFIASSSTLSKHEIHTGCKP 267
 QY 100 FVCAATG-----CDOKNTFSNLKHEERKENOQKOYI 133
 Db 268 YKCECGKAFNOSSTLTLLKHKIHTGEKPYKCECGKAFNOSSTLTLLK--KKIHTGEKPY 325
 QY 134 CSPEDCKRTTYFNLQSHILSFHESSRPVCEHAGCGCTAMKOSLTRNAVNDPDKK--YV 190
 Db 326 C--EBCGKAFVASSSTLSKHEI--IHTGCKPYKCE--CGKAFIASSSTLSHFEIHTGCKHYK 381
 QY 191 CGK--GCSFVAKTMTLLKHNRETHKEE--ILCEVCRKPFKRDYLLQHKMTHAPERDVC 247
 Db 382 CEECGKAFI---WSSVLTTRHKRYHTGEKPYKCECGKAFKYSSTLSHKSHTGEKPY-K 437
 QY 248 CPRGCGRTTYTFNLQSHILSFHESSRPVCEHAGCGCTAMKOSLTRNAVNDPDKK- 306
 Db 438 C--EBCGKAFVASSSTLSKHEI--IHTGCKPYKCE--CGKAFIASSSTLSKHEIHTGCKPY 492
 QY 307 KMKLVKRSREKRSLSHLSGYLPPRKOGGSLCQN-----GESPVCYEDKMLST 358
 Db 493 KCECGKAFNOSSTLTLLKHKIHTGEKPYKCECGKAFNOSSTLTLLKHKIHT 543

RESULT 13
 ABG18386
 ID ABG18386 standard; Protein; 632 AA.
 XX
 AC ABG18386;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #18377.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI: 2001-639362/73.
 DR N-PSDB: AAS82573.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 20; SEQ ID NO 48745; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30327 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 632 AA:
 Query Match 20.4%; Score 410.5; DB 22; Length 632;
 Best Local Similarity 30.5%; Pred. No. 9.1e-26;
 Matches 109; Conservative 48; Mismatches 137; Indels 63; Gaps 17;
 Db 269 YKCECGKAFNOSSTLTLLKHKIHTGEKPYKCECGKAFNOSSTLTLLK--KKIHTGEKPY 326
 QY 134 CSPEDCKRTTYFNLQSHILSFHESSRPVCEHAGCGCTAMKOSLTRNAVNDPDKK--YV 190
 Db 327 C--EBCGKAFVASSSTLSKHEI--IHTGCKPYKCE--CGKAFIASSSTLSHFEIHTGCKHYK 382
 QY 191 CGK--GCSFVAKTMTLLKHNRETHKEE--ILCEVCRKPFKRDYLLQHKMTHAPERDVC 247
 Db 383 CEECGKAFI---WSSVLTTRHKRYHTGEKPYKCECGKAFKYSSTLSHKSHTGEKPY-K 438
 QY 248 CPRGCGRTTYTFNLQSHILSFHESSRPVCEHAGCGCTAMKOSLTRNAVNDPDKK- 306
 Db 439 C--EBCGKAFVASSSTLSKHEI--IHTGCKPYKCE--CGKAFNOSSTLTLLKHKIHTGCKPY 493
 QY 307 KMKLVKRSREKRSLSHLSGYLPPRKOGGSLCQN-----GESPVCYEDKMLST 358
 Db 494 KCECGKAFNOSSTLTLLKHKIHTGEKPYKCECGKAFNOSSTLTLLKHKIHT 544

RESULT 14
 AAM79171
 ID AAM79171 standard; Protein; 652 AA.
 XX
 AC AAM79171;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1833.
 KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukemia;
 XX nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 PN WO200157190-A2.
 PD 09-AUG-2001.
 PF 05-FEB-2001; 2001WO-US04098.
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.

XX		PA	(HYSE-) HYSED INC.	
PR	30-NOV-2000:	2000US-0728422.		
PI	Tang YT,	Liu C,	Dzianac RT,	Aasund V,
PI	Zhang QA,	Wang D,	Wang J,	Zhang J,
PI	Xue AJ,	Yang Y,	Wejhrman T,	Goodrich R;
DR	WPL: 2001-476283/51.			
DR	N-PDSB: AAK52304.			
PT	Nucleic acids encoding polypeptides with cytokine-like activities,			
PT	useful in diagnosis and gene therapy -			
PS	Claim 20; page 4219-4221: 6221pp; English.			
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the			
CC	encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to			
CC	cytokine, cell proliferation or cell differentiation or which may induce			
CC	production of other cytokines in other cell populations. The			
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or			
CC	peptide therapy. The polypeptides have various cytokine-like activities,			
CC	e.g. stem cell growth factor activity, haematopoiesis regulating			
CC	activity, tissue growth factor activity, immunomodulatory activity and			
CC	activin/inhibin activity and may be useful in the diagnosis and/or			
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and			
CC	Inflammation.			
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666			
CC	(AAM80020) are omitted as the relevant pages from the sequence listing			
CC	were missing at the time of publication.			
XX	Sequence	652 AA;		
SQ				
Query Match		20.4%;	Score 410.5;	DB 22;
Best Local Similarity		37.7%;	Pred. No. 9.5e-26;	
Matches	100;	Conservative	31;	Mismatches 111;
				Indels 23;
				Gaps 14;
OY	46	DCSANNISKAMKDANLCIKTGERPRVCDEEGCGKAFLINDYLSNHLITHTGKERPVCAAT	105	
DB	375	ECSGAFFWSSSLNKHNRITHTGEKPCTC--EECGKAFRSHSLANKRKIRHTGEEKPYCE-	431	
OY	106	GCDQKTEKSNLKKHFERKNENQKOYLICSPEDCKTKPKKHQQCLKINOCOTNPFLPCT	165	
DB	432	-CGKAFNOSTLLIH-KRIHSG-QRPYC--EECGKAFTSTLTNEHKKIHTGEEKPYKC-	485	
OY	166	QECCGKFAPASRLSKRNAKANHS---YQCQSCSVAVATTVELLKRNRETNKEIL--CE	220	
DB	486	-EFCGKAFIKSALNLNHNKNIHTGEEKPYKC-EECGKAFOSSGLTIH-RSIHSEQKLKCE	542	
OY	221	VCRKTERKDYLNQHMKKTAPARPDVRCSPREGCGRTTYTVNNLOSHLSPHEESRPVCE	280	
DB	543	ECSGAFFRSTALNHNKIIHGSEKPY-KC--KECGKAYULSTLIKH-KRIHTGEEKPTCE	598	
OY	281	HAGCGKTEFAKOSTLRNAVVDPRK	305	
DB	599	E--CGKAFWKSSSLTKNHKITHTGEC	621	
RESULT 15				
ID	ABB64531	standard; Protein: 403 AA.		
AC	ABB64531;			
DT	26-MAR-2002	(first entry)		
DE	Drosophila melanogaster polypeptide SEQ ID NO 20385.			
KM	Drosophila; developmental biology; cell signalling; insecticide;			
OS	Drosophila melanogaster.			

```

XX MN 2000171042-A2.
PD XX 27-SEP-2001.
PP 23-MAR-2001; 2001MO-US09231I.
PX XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
PX XX (PEKE ) PE CORP NY.
PA Venter JC, Adams M, LI PWD, MYERS EW:
PI MPI; 2001-656860/75.
DR N-PDB; ABL08634.
XX XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions .
PS Disclosure; SEQ ID NO 20385; 21pp + Sequence Listing: English.
XX XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABR01840-ABL16175) and the encoded proteins
CC (ABB57737-ABR72072).
CC CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 403 AA:

Query Match 20.2% Score 406.5 DB 22 Length 403:
Best Local Similarity 30.3% Pred. No. 1.1e-25:
Matches 105; Conservative 55; Mismatches 144; Indels 43; Gaps 14:

QY 39 RFLCSFPCDSANYSKAMKLDHLNCKNGSERPVCDYEGSCAFIRDYHLSRI-I-LTH-- 94
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 KYACSMKFCATGTRKLDLDLRHEYNHTGIKKMKASCYECDTYIVTLKRHSRTHERP 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 --TGEKRVCAALAGCDOKENTSNLKIKREKNENQOYICSEDCDCTFKKKNOOLKH 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 94 ESAAKTKVVCALCEGCKMFTSVSNATGRMYRETNES-KPVYPCS--QCSAASFQSLKLKH 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 153 QC-QHNTERPLFKCTDGCGCNAPASPKLYRNAKAHEGVYCGSGCFYAKTWETYLKNVE 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 EIREHTLEUYPCSS--CSRGFYQOMOCOSHSPRSCKILEC-RGCRPLDFDKMTLYTKCHRD 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 T-----HKELLCEVCRTEFKRKLYLOLNHIKT---NARENDVC----RCRRGGCGRTYT 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 SLNEGKRNHRK---CDRCDSAFKPSELRLNRHLEVUKHKEAAQUDGCATGFTNEGCGCKSTVS 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 259 TVENLOSILTSFEHESRPVCEHACCQGTFAFKOSITLTGHAV-VIDPRKMMKLYVKKSRE 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 YLRNLRLQHMLTA-SGRRFECQALDSCGCFSSAQQLANHLRYLDKDAQTKELKAKKKOK 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 318 KRSLASHLSGYIPRKKGOGGLSLCOGNESPVCEDKMJSLVANVLTL 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 SKTGGEGKTKSTRKKRRRDGSR-----KHSRLSKLAQLDL 358

```

Search completed: February 10, 2003, 17:45:15
Job time : 77 secs

1
2
3
4

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:44:27 : Search time 24 Seconds
(without alignments)
447.474 Million cell updates/sec

Title: US-09-831-426c-2

Perfect score: 2008

Sequence: 1 MDPPAVVAESVSLTIADAF.....ESPNCVEDKMLSTVAVLTLG 365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database : Issued.Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/CTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1933	96.3	423	1	US-08-523-376-3
2	381	19.0	711	2	US-08-820-170A-10
3	381	19.0	711	3	US-09-055-699-10
4	381	19.0	711	4	US-09-273-565-10
5	381	19.0	711	4	US-09-565-538-10
6	381	19.0	711	4	US-09-661-468-10
7	353.5	17.6	543	4	US-09-562-123A-4
8	336	16.7	803	4	US-09-063-035-2
9	318.5	15.9	727	4	US-08-475-844-9
10	318.5	15.9	727	5	PCT-US95-08429-9
11	313	15.6	640	4	US-09-262-773-4
12	313	15.6	648	4	US-09-262-773-2
13	312.5	15.6	462	3	US-08-486-099-117
14	312.5	15.6	462	3	US-08-560-107A-127
15	312.5	15.6	462	3	US-08-484-223B-117
16	312.5	15.6	462	3	US-08-919-597-117
17	312.5	15.6	462	3	US-08-475-668A-117
18	312.5	15.6	462	3	US-08-485-551A-117
19	312.5	15.6	462	3	US-08-471-913A-117
20	312.5	15.6	462	4	US-08-485-264A-117
21	312.5	15.6	462	4	US-08-474-349A-117
22	312	15.5	728	2	US-08-475-844-5
23	312	15.5	728	5	PCT-US95-08429-5
24	301.5	15.0	488	2	US-08-933-750C-17
25	301.5	15.0	488	4	US-09-234-613-17
26	288.5	14.4	181	4	US-08-676-318A-44
27	278.5	13.9	181	4	US-08-863-813A-44

28	259.5	12.9	667	2	US-08-718-661-2	Sequence 2, Appl1
29	251.5	12.5	441	4	US-09-172-045-2	Sequence 2, Appl1
30	248	12.4	345	1	US-08-102-942A-2	Sequence 2, Appl1
31	248	12.4	345	4	US-09-037-179B-2	Sequence 2, Appl1
32	246	12.3	429	1	US-08-234-783-4	Sequence 4, Appl1
33	246	12.3	429	1	US-08-456-907-4	Sequence 4, Appl1
34	246	12.3	429	5	PCT-US95-05523-4	Sequence 4, Appl1
35	243.5	12.1	449	1	US-08-102-942A-4	Sequence 4, Appl1
36	243.5	12.1	449	4	US-09-037-179B-4	Sequence 4, Appl1
37	242.5	12.1	447	3	US-09-234-332-11	Sequence 11, Appl1
38	242.5	12.1	449	1	US-08-102-942A-6	Sequence 6, Appl1
39	242.5	12.1	449	4	US-09-037-179B-6	Sequence 6, Appl1
40	238	11.9	338	2	US-08-933-750C-4	Sequence 4, Appl1
41	238	11.9	338	4	US-09-234-613-4	Sequence 4, Appl1
42	224.5	11.2	671	3	US-09-121-321-16	Sequence 16, Appl1
43	224.5	11.2	671	4	US-08-933-803A-16	Sequence 16, Appl1
44	222	11.1	414	5	PCT-US92-06840-2	Sequence 2, Appl1
45	221	11.0	706	1	US-08-074-967-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-08-523-376-3
Sequence 3, Application US/08523376
Patent No. 568030
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Satoshi, TAKEEDA
APPLICANT: Yoshikazu, SHIMADA
APPLICANT: Kouichi, OZAKI
APPLICANT: Sadehito, SIN
TITLE OF INVENTION: HPTIIA GENE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESS: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3302
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,376
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-523-376-3

Query Match 96.3%: Score 1933; DB 1; Length 423;
Best Local Similarity 96.2%: Pred. No. 8.3e-189;
Matches 351; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MDPPAVVAESVSLTIADAFIAGESSAPPPALPRRIFCSPPDCSANNKAKKLDAAH 60
DB 59 LDPPAVVAESVSLTIADAFIAGESSAPPPALPRRIFCSPPDCSANNKAKKLDAAH 118
QY 61 LCKHNGEPPFCDEGCGKAPITRDYHLSRIILHTGEKPPVCAAGCCDDKPFKSNLKKH 120
|||||
```

Db 119 LCKHTGERPFVCOYEGCGKAFIRDYHLNRIHLTHTEGKRPVCAANGCQCKQENTKSNLKKH 178
Qy 121 FERKHEMOQOYICSPEDCKKTRKKQOJ.KIHQCOHTNEPLFKTQEOCCGHNFPSPK 180
Db 179 FERKHEMOQOYICSPEDCKKTRKKQOJ.KIHQCOHTNEPLFKTQEOCCGHNFPSPK 238
Qy 181 RHAKAHGVCYQKSCSFVAKTWTELLKHVRETHKEBILCEVCRKTRFKKQJLKHMTA 240
Db 239 RHAKAHGVCYQKSCSFVAKTWTELLKHVRETHKEBILCEVCRKTRFKKQJLKHMTA 298
Qy 241 PERDYCRCPRECGGRTTYTVFNQSHILSFHEESRPVCEHAGCGKTRAMQSLTRAVV 300
Db 299 PERDYCRCPRECGGRTTYTVFNQSHILSFHEESRPVCEHAGCGKTRAMQSLTRAVV 358
Qy 301 HDPKXKKMKLKVKKSRKREKSLASHLSGYIPPKRQOGLSLQNGESPVCYEDKMLSTVA 360
Db 359 HDPKXKKMKLKVKKSRKREKSLASHLSGYIPPKRQOGLSLQNGESPVCYEDKMLSTVA 418
Qy 361 VLTLG 365
Db 419 VLTLG 423

RESULT 2

US-08-820-170A-10
; Sequence 10, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-820-170A-10

Query Match 19.0%; Score 381; DB 2; Length 711;
Best Local Similarity 32.9%; Pred. No. 2,6e-30;
Matches 93; Conservative 43; Mismatches 119; Indels 28; Gaps 14;

Qy 38 RRFICSPDCSANSKAMKLDALHLCNTEGPRFVCDYEGCGKAFIRDYHLNRIHLTHTEG 97
Db 445 KSYVCI--EGGQAFIQAHLIVHQRSHTEGKRPYC--HNCGKSFISKQDLINRIHTGE 500
Qy 98 KRPVCAATGCDQKFNTPSNLKKHFERKHEMOQOYICSPEDCKKTRKKQJLKHQCOHT 157
Db 501 KPYVCS--SDCGKTFQKSHLNTH--QKHTGERHNVC--EGGKAFNCKSLTSMHQJHT 554

Qy 158 NEPLFKTQEOCCGHNFPSPKLRHAKAHG---VYQKSCSFVAKTWTELLKH-VRETH 213
Db 555 GEKPYKCS--CGKAFISKQFKEHQRHTTGERKPYVCE--CGKAFNGRSHNKHQJHTTR 611
Qy 214 KEELICEVCRKTRFKKQJLKHMTAHAPERDYCRPRE--CGGRTTYTVFNQSHILSFH 271
Db 612 ERPPVCKKCGAFQKSELLTHQRTHMER-----PYECLDQKSFSSKPOLKYN-QRIH 665
Qy 272 EESRPVCEHAGCGKTRAMQSLTRHAVVHDPDK-KMKLYK 313
Db 666 TGERPYVCS--CGKAFNKNKSNFKHQTHTTRDKSYKSYSYK 706

RESULT 3

US-09-055-699-10
; Sequence 10, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION: *
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-055-699-10

Query Match 19.0%; Score 381; DB 3; Length 711;
Best Local Similarity 32.9%; Pred. No. 2,6e-30;
Matches 93; Conservative 43; Mismatches 119; Indels 28; Gaps 14;

Qy 38 RRFICSPDCSANSKAMKLDALHLCNTEGPRFVCDYEGCGKAFIRDYHLNRIHLTHTEG 97
Db 445 KSYVCI--EGGQAFIQAHLIVHQRSHTEGKRPYC--HNCGKSFISKQDLINRIHTGE 500
Qy 98 KRPVCAATGCDQKFNTPSNLKKHFERKHEMOQOYICSPEDCKKTRKKQJLKHQCOHT 157
Db 501 KPYVCS--SDCGKTFQKSHLNTH--QKHTGERHNVC--EGGKAFNCKSLTSMHQJHT 554
Qy 158 NEPLFKTQEOCCGHNFPSPKLRHAKAHG---VYQKSCSFVAKTWTELLKH-VRETH 213
Db 555 GEKPYKCS--CGKAFISKQFKEHQRHTTGERKPYVCE--CGKAFNGRSHNKHQJHTTR 611
Qy 214 KEELICEVCRKTRFKKQJLKHMTAHAPERDYCRPRE--CGGRTTYTVFNQSHILSFH 271

```
Db      612 ERPVCTCKCGKAPVQKSELTHRTHTMGK-----PYECLDGGKFSKKPOLKVN-QRIH 665
Qy      272 ESRPFVCEHAGCGKTFAMKQSLTRHAVVHDPDK-KMKLYK 313
Db      666 TGERPVYCS--CGKAFNNSNPNKHQTHTRDKSKYSYSVK 706

RESULT 4
US-09-273-565-10
: Sequence 10, Application US/09273565A
: Patent No. 6166190
: GENERAL INFORMATION:
: APPLICANT: FUJIMURA, TSUTOMU
: APPLICANT: MATANABE, TAKESHI
: APPLICANT: HORIE, MASATO
: TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
: TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
: FILE REFERENCE: Q-53599
: CURRENT APPLICATION NUMBER: US/09/273,565A
: EARLIER FILING DATE: 1999-03-22
: EARLIER APPLICATION NUMBER: 09/055,699
: EARLIER FILING DATE: 1998-04-07
: EARLIER APPLICATION NUMBER: 08/820,170
: EARLIER FILING DATE: 1997-03-19
: EARLIER APPLICATION NUMBER: JP 63410/1996
: EARLIER FILING DATE: 1996-03-19
: EARLIER APPLICATION NUMBER: JP 69163/1997
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 10
: LENGTH: 711
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-273-565-10

Query Match      19.0%; Score 381; DB 4; Length 711;
Best Local Similarity 32.9%; Pred. No. 2,6e-30;
Matches 93; Conservative 43; Mismatches 119; Indels 28; Gaps 14;

Qy      38 RRFICSPDCSANTSKMKIDANLCKNTGERRPVCDYECGCAFTIDYLSNHLITGGE 97
Db      445 KSYVCI--EGQAFIQKANHILVHQRSHTEKRPYC--HNCGSFISKSQOLDIHRHITGE 500
Qy      98 KPFVCAATGCDOKENTSNLKKHFERKHNOCQOYICSPEDCKTFKKHQDLKHOCOUT 157
Db      501 KPYEC--SDGKFTPTQKSHLNH--QKIHGGERHNVC--ECGKAFNCKSLTSMHQRIHT 554
Qy      158 NEPLFKTQEGCGKHNFSKSLKRYAKANDG---YVCQKGSFVAKTTELLKH-VRETH 213
Db      555 GEPKPKCE--CGKAFISKSQFKEHORIHTGEPKPYCTE--CGKAFNGSNFNKHQDITHTR 611
Qy      214 KEELICEVCKRTFKRKDYLLKQNMKTHAPRDVCRPRE--GCGRTYTVFMLOSHILSFH 271
Db      612 ERPFVCKCGKAPVQKSELTHRTHTMGK-----PYECLDGGKFSKKPOLKVN-QRIH 665
Qy      272 ESRPFVCEHAGCGKTFAMKQSLTRHAVVHDPDK-KMKLYK 313
Db      666 TGERPVYCS--CGKAFNNSNPNKHQTHTRDKSKYSYSVK 706

RESULT 5
US-09-565-538-10
: Sequence 10, Application US/09565538
: Patent No. 6333404
: GENERAL INFORMATION:
: APPLICANT: FUJIMURA, TSUTOMU
: APPLICANT: WATANABE, TAKESHI
: APPLICANT: HORIE, MASATO
: TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
: TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
: FILE REFERENCE: Q-53599
```

```
Qy      38 RRFICSPDCSANTSKMKIDANLCKNTGERRPVCDYECGCAFTIDYLSNHLITGGE 97
Db      445 KSYVCI--EGQAFIQKANHILVHQRSHTEKRPYC--HNCGSFISKSQOLDIHRHITGE 500
Qy      98 KPFVCAATGCDOKENTSNLKKHFERKHNOCQOYICSPEDCKTFKKHQDLKHOCOUT 157
Db      501 KPYEC--SDGKFTPTQKSHLNH--QKIHGGERHNVC--ECGKAFNCKSLTSMHQRIHT 554
Qy      158 NEPLFKTQEGCGKHNFSKSLKRYAKANDG---YVCQKGSFVAKTTELLKH-VRETH 213
Db      555 GEPKPKCE--CGKAFISKSQFKEHORIHTGEPKPYCTE--CGKAFNGSNFNKHQDITHTR 611
Qy      214 KEELICEVCKRTFKRKDYLLKQNMKTHAPRDVCRPRE--GCGRTYTVFMLOSHILSFH 271
Db      612 ERPFVCKCGKAPVQKSELTHRTHTMGK-----PYECLDGGKFSKKPOLKVN-QRIH 665
Qy      272 ESRPFVCEHAGCGKTFAMKQSLTRHAVVHDPDK-KMKLYK 313
Db      666 TGERPVYCS--CGKAFNNSNPNKHQTHTRDKSKYSYSVK 706

RESULT 6
US-09-661-468-10
: Sequence 10, Application US/09661468
: Patent No. 6376189
: GENERAL INFORMATION:
: APPLICANT: FUJIMURA, TSUTOMU
: APPLICANT: WATANABE, TAKESHI
: APPLICANT: HORIE, MASATO
: TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
: TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
: FILE REFERENCE: Q-53599
: CURRENT APPLICATION NUMBER: US/09/661,468
: PRIOR FILING DATE: 2000-09-13
: PRIOR APPLICATION NUMBER: 09/055,699
: PRIOR FILING DATE: 1998-04-07
: PRIOR APPLICATION NUMBER: 08/820,170
: PRIOR FILING DATE: 1997-03-19
: PRIOR APPLICATION NUMBER: JP 63410/1996
: PRIOR FILING DATE: 1996-03-19
: PRIOR APPLICATION NUMBER: JP 69163/1997
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 10
: LENGTH: 711
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-661-468-10
```

Query Match	19.0%	Score 301;	DB 4;	Length 711.
Best Local Similarity	32.9%	Pred. No. 2,66-30;		
Matches	99;	Conservative	43;	Mismatches 119; Indels 28; Gaps 14.

Qy	38	RRTGCSPPDCSANTSKAMKIDAHLLKNTGCERPVCDVCGCGAFIRDYHLSRHILTHTGE	97
Db	445	KSYVCI--ECGAFQIKAMHLYQNRSHGTGEKYUOC--HNGCKSEIYSOLDINRHITGE	500
Qy	98	KPEVCATGCGDQGFNTKRSNLKHFENKHEENQOYQISFCSECKSTFFKNOOLKHOCQH	157
Db	501	KPYDC--SDCGKTFEQSHLNTI--QKIHGTGRHHVCS--ECGAFQPKSLTSHNQKIH	554
Qy	158	NEPFLKTCGSCCKHNAPSRKLYAKNAKAE---VVCQKGSFPAKWTIELKH-VRETH	213
Db	555	GKPYKQSE--CGKATSKSOFREHORITGKPRVCE--CGKAFNGRSNPNKHQITHTH	611
Qy	214	KEELICEVGRKTERRDYLQNHKMTTHAREDVCKCPRE--CGGRTTYTNNLSHLSH	271
Db	612	ERPVCYCKCKAEVQKSELITHQRTHMGEK----PYCLSDCGKSPSKPOLKVN--QRH	665
Qy	272	EESRPVCEHAGCGKTEFAKQKSLTRAVVHDDCK--KKKLTVK	313
Db	666	TGERPYVCSF--CGKAFNNRSNPNKHQITHTHDKSYKSYVK	706

```

RESULT 7
US-09-362-123A-4
: Sequence 4, Application US/09362123A
: Patent No. 6451558
: GENERAL INFORMATION:
: APPLICANT: Cooke, Michael Paul
: APPLICANT: Holness, Claire
: APPLICANT: Sirenko, Oksana
: TITLE OF INVENTION: No. 6451558el Genes in the Control of Hemttopoiesis
: FILE REFERENCE: 4-306294/SVS
: CURRENT APPLICATION NUMBER: US/09/362,123A
: CURRENT FILING DATE: 1999-07-27
: PRIORITY APPLICATION NUMBER: 09/128,310
: PRIOR FILING DATE: 1998-08-03
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 543
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-362-123A-4

```

Query Match	17.68;	Score 353.5;	DB 4;	Length 543;
Best Local Similarity	31.58;	Pred. No. 1.2e-27;		
Matches 100; Conservative	38;	Mismatches 144;	Inbels 35;	Gaps 15;

[illegible]

```

Db      511 RSVAGASSEGAPAKETE 527
          ||:| | | :

```

RESULT B
 US-09-063-035-2
 Sequence 2. Application US/09063035
 Patent No. 6160091
 GENERAL INFORMATION:
 APPLICANT: PEUKERT, Karen; HAENEL, Frank; and EILERS,
 APPLICANT: Martin
 TITLE OF INVENTION: Myc-binding zinc finger proteins,
 TITLE OF INVENTION: their preparation and their use
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Kell & MelnkauF
 STREET: 1101 Connecticut Avenue
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: diskette, 3.5 inch, 1.2 Mb storage
 COMPUTER: IBM AT-compatible, 80486 processor
 OPERATING SYSTEM: MS-DOS version 6.1
 SOFTWARE: Wordperfect version 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/063.035
 FILING DATE: 21-APR-1998
 CLASSIFICATION: 514
 INFORMATION FOR SEQ. ID NO. 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 803 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-09-063-035-2

Query Match	16.7%;	Score 336;	DB 4;	Length 803;
Best Local Similarity	31.8%;	Pred. No. 1.2e-25;		
Matches	95;	Conservative	45;	Mismatches 117;
				Indels 42;
				Gaps 16;

[illegible]

RESULT 9
US-08-475-844-9
Sequence 9
Application US/08475844
Patent No. 5972643
GENERAL INFORMATION:
APPLICANT: Lohamkov, Victor V
APPLICANT: Nelman, Paul E.
APPLICANT: Klenova, Elena M.
APPLICANT: Goodman, Graham H.
APPLICANT: Filipova, Galina N.
APPLICANT: Collins, Steven J.


```

: TITLE OF INVENTION: CTCF
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend Kourie and Crew
: STREET: One Market Plaza
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94105
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,844
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 536
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: 08/261,680
: FILING DATE: 17-JUN-1994
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Parmelee, Steven W.
: REGISTRATION NUMBER: 31,990
: REFERENCE/DOCKET NUMBER: 14538A-11-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-467-9600
: TELEFAX: 415-543-5043
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 727 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-475-844-9

```

Query Match 15.9%; Score 318.5; DB 2; Length 727;
Best Local Similarity 27.6%; Pred. No. 6.3e-24;
Matches 103; Conservative 41; Mismatches 136; Indels 93; Gaps 18;

```

QY 31 PPPR-----ALPRRFISPPDCSANSKAMKLDALHCNHTGERPFVCDYEGCGKAFTRD 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 PPKPRFKIKKKVKKTFQCEL--CSYCPRRSNLDRMKSHTERDPHKCHL--CGRAFPTV 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 85 YLHSNHLITHTGERPFVCAATGCDOKFNTKSNLKKHFERRKHNQOKYICSEFDCKTFK 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 TLLRNHLNTHGTGRPHKC--PDCDMAFVTSGELVRRRYKH--THEKPRCSM--CDYASV 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145 KHQQLKIHQCOHTNEPLFKCTOEGCGKNFASPSKLRNKAHNG-----YVC-----QK 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 EYSKLRKRIIRSHTGERPFQCSL--CSYASRDTYKLRKRNHRTSHGKPEYECYICARFTQS 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 194 G-----CSFYAKTWTLLKLVNRETH--KEELICEVCSEKTFPKK 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 GTMKNHILQKTEVNAKFNCHPCDVTIARKSDLGVLHROHSYIRQGGKRCRYCDVAFHER 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 230 DYLKOHMKNTHAPER---DYC--RCPRE-----CGGRTYTVVFNQ 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 YALLIOHOKSHNEKRFKCDODYACROERIMIMKRTHTGKPYACSHCDKTFROKOLD 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 265 SHLSFHEES---RPVCEHAGCGKTFAMKOSLTRHA--VHNDP-----KKMKL 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 MHFKRYHDPNVPAFVVC--SKCGKTFTRNTMARHADNACAPDVGEGENGGETKKSRG 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 311 KVKKSREKRSLAS 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 KKKMKRSKEDSS 610

```

```

: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: CTCF
: NUMBER OF SEQUENCES: 21
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/08429
: FILING DATE: 15-JUN-1995
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/261,680
: FILING DATE: 17-JUN-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Parmelee, Steven W.
: REGISTRATION NUMBER: 31,990
: REFERENCE/DOCKET NUMBER: 14538A-11-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-467-9600
: TELEFAX: 415-543-5043
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 727 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US95-08429-9

```

Query Match 15.9%; Score 318.5; DB 5; Length 727;
Best Local Similarity 27.6%; Pred. No. 6.3e-24;
Matches 103; Conservative 41; Mismatches 136; Indels 93; Gaps 18;

```

QY 31 PPPR-----ALPRRFISPPDCSANSKAMKLDALHCNHTGERPFVCDYEGCGKAFTRD 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 PPKPRFKIKKKVKKTFQCEL--CSYCPRRSNLDRMKSHTERDPHKCHL--CGRAFPTV 306
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 85 YLHSNHLITHTGERPFVCAATGCDOKFNTKSNLKKHFERRKHNQOKYICSEFDCKTFK 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 TLLRNHLNTHGTGRPHKC--PDCDMAFVTSGELVRRRYKH--THEKPRCSM--CDYASV 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 145 KHQQLKIHQCOHTNEPLFKCTOEGCGKNFASPSKLRNKAHNG-----YVC-----QK 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 EYSKLRKRIIRSHTGERPFQCSL--CSYASRDTYKLRKRNHRTSHGKPEYECYICARFTQS 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 194 G-----CSFYAKTWTLLKLVNRETH--KEELICEVCSEKTFPKK 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 GTMKNHILQKTEVNAKFNCHPCDVTIARKSDLGVLHROHSYIRQGGKRCRYCDVAFHER 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 230 DYLKOHMKNTHAPER---DYC--RCPRE-----CGGRTYTVVFNQ 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 YALLIOHOKSHNEKRFKCDODYACROERIMIMKRTHTGKPYACSHCDKTFROKOLD 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 265 SHLSFHEES---RPVCEHAGCGKTFAMKOSLTRHA--VHNDP-----KKMKL 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 MHFKRYHDPNVPAFVVC--SKCGKTFTRNTMARHADNACAPDVGEGENGGETKKSRG 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 311 KVKKSREKRSLAS 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 KKKMKRSKEDSS 610

```

RESULT 11
US-09-262-773-4
; Sequence 4, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne

[illegible][illegible][illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:45:22 : Search time 18 Seconds
(without alignments)
449.544 Million cell updates/sec

Title: US-09-831-426c-2
Perfect score: 2008
Sequence: 1 MDPFAVAVESVSLRIADAF.....ESPNCVEDKMLSTVAVLTLG 365

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 129505 seqs, 22169297 residues

Total number of hits satisfying chosen parameters: 129505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCIT_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/PCITUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	386	19.2	406	10	US-09-764-864-976 Sequence 976, App
2	383.5	19.1	458	10	US-09-764-864-909 Sequence 909, App
3	383.5	19.1	457	10	US-09-764-864-1045 Sequence 1045, Ap
4	381	19.0	711	10	US-09-976-165-10 Sequence 10, Appl
5	381	19.0	711	10	US-09-828-648-2 Sequence 2, Appl
6	379	18.9	525	10	US-09-864-761-33653 Sequence 33653, A
7	375	18.7	338	10	US-09-764-864-948 Sequence 948, App
8	372	18.5	551	10	US-09-864-761-37065 Sequence 37065, A
9	368.5	18.4	515	10	US-09-864-761-37152 Sequence 37152, A
10	366.5	18.3	278	10	US-09-864-761-42958 Sequence 42958, A
11	366.5	18.3	516	9	US-09-974-298-118 Sequence 118, App
12	366.5	18.3	547	10	US-09-764-864-1065 Sequence 1065, Ap
13	366	18.2	610	9	US-09-989-920-212 Sequence 212, App
14	360.5	18.0	580	10	US-09-925-301-1357 Sequence 1357, Ap
15	359	17.9	186	9	US-09-942-087A-30 Sequence 30, Appl
16	359	17.9	156	10	US-09-779-233-18 Sequence 29, Appl
17	359	17.9	156	10	US-09-844-508-29 Sequence 18, Appl
18	356.5	17.8	468	10	US-09-864-761-37991 Sequence 37991, A
19	356.5	17.8	536	10	US-09-864-761-36148 Sequence 36148, A

20	355.5	17.7	241	10	US-09-764-864-935 Sequence 935, App
21	355.5	17.7	492	10	US-09-864-761-42897 Sequence 42897, A
22	355.5	17.7	605	10	US-09-864-761-36132 Sequence 36132, A
23	355.5	17.7	605	10	US-09-864-761-37943 Sequence 37943, A
24	354	17.6	434	10	US-09-864-761-33476 Sequence 33476, A
25	352.5	17.6	513	10	US-09-764-864-1061 Sequence 1061, Ap
26	351	17.5	269	10	US-09-864-761-37651 Sequence 37651, A
27	349.5	17.4	269	10	US-09-864-864-887 Sequence 887, App
28	349.5	17.4	284	10	US-09-764-864-932 Sequence 932, App
29	347.5	17.3	558	10	US-09-864-761-47724 Sequence 47724, A
30	347.5	17.3	585	10	US-09-764-864-1063 Sequence 1063, Ap
31	344	17.1	340	10	US-09-864-761-35894 Sequence 35894, A
32	343.5	17.1	361	10	US-09-764-864-893 Sequence 893, App
33	342.5	17.1	483	10	US-09-764-864-940 Sequence 940, App
34	342.5	17.1	614	10	US-09-864-761-43107 Sequence 43107, A
35	340.5	17.0	311	10	US-09-864-761-34333 Sequence 34333, A
36	336.5	16.8	413	10	US-09-864-761-37010 Sequence 37010, A
37	334	16.6	252	10	US-09-816-669A-12 Sequence 12, Appl
38	331.5	16.5	333	10	US-09-764-864-858 Sequence 958, App
39	331	16.5	577	10	US-09-764-864-883 Sequence 883, App
40	330.5	16.5	257	10	US-09-764-864-1428 Sequence 1428, Ap
41	327.5	16.3	254	10	US-09-864-761-33477 Sequence 33477, A
42	326	16.2	375	10	US-09-864-761-36336 Sequence 36336, A
43	325.5	16.2	281	10	US-09-764-864-1076 Sequence 1076, Ap
44	324	16.1	430	10	US-09-864-761-37994 Sequence 37994, A
45	323.5	16.1	485	10	US-09-864-761-37026 Sequence 37026, A

ALIGNMENTS

RESULT 1

US-09-764-864-976
Sequence 976, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT23
CURRENT FILING DATE: 2001-01-17
CURRENT APPLICATION NUMBER: US/09/764,864
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 976
LENGTH: 406
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-976

Query Match	Best Local Similarity	Score	Pred. No.	DB	Length
Matches 105; Conservative	33.9%	386;	8e-25;	10;	406;
			Matches 125;	Indels 42;	Gaps 17;
47	CGANYSKAMKIDAHLCNHTGEPVCDYEGCGKAFIRDYNIYRNILYHTGKRPVCAATG	106			
82	CSMNSHLLMPQKS---HTGKRPYC--RPGCAAFSEKSRKLNHORTHGKRPYC--DG	133			
107	CDQKPNYSNLKHPREKNEQKQYICSEFDCKKFPKKHQDLKHOSQHTPEPFTQ	166			
134	CDKAFSAVSGLRTH-QRTHG-EKRPFC--HEGGSFNYKSLIYQIRHTGKRFEDNE	189			
167	EGCGKHFSPKRLKRNKAHEG--UYCQKSGCFVAKTTELLKVR-ETNHEELICEVC	222			
190	--CGKSFHMSGRLNHRHTHTGGERPYCDE--CKAKFKLSGLRKHNRYTGKRPYKQNC	246			
223	KRTFKRKYLLKQNHKTAPREDYRCRPRECCGRTYTVTNLQSHLSFNEERPVCSHA	282			
247	GAAGQKQSLRGHNHTHTGKRPY-KC--NHGGAFOKSNLRYNHT-HTGKRPQCE--	301			
283	CGGKTFPAKQKSLTRNAVVDPPDKKMKLKYS--REKRSIASHLSGYIPPKKQGGSL	341			
302	--CGKTFROKSNLRYNHTHTGKRPYECNECGKAFSEKSVLRKH-----OR	345			

OY 342 CONGESP-NC 350
Db 346 THTEGKPYNC 355

RESULT 2

US-09-764-864-909
: Sequence 909, Application US/09764864
: Patent No. US20020132753A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT223
: CURRENT APPLICATION NUMBER: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1792
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 909
: LENGTH: 448
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-764-864-909

Query Match 19.1%; Score 383.5; DB 10; Length 448;
Best Local Similarity 34.0%; Pred. No. 1.8e-24;

Matches 115; Conservative 40; Mismatches 134; Indels 49; Gaps 22;

OY 42 CSFPDCCSANYSKAMKLDALHCKHTGERPVCDEGCGKAFIRDYHLSRHLTHTEGKPF 101
Db 35 CSV--CGKARSTKFSLTTHQKHTGKPYIC--SECCKGFIKRLLIHNRHTHTGKPF 90
OY 102 CAATGCCOKENTSNLKKHFERKHEHQKQYICSPEDCKKTFKKHQDLKIHQCOHTNEPL 161
Db 91 C--KKCKGKGLTKNSLTH--QDTHG--EKLYTCS--EGCGKFSMKHCLMVAHQHTGKEXP 144
OY 162 FKCTQEGSCGHFASPSKLRHAKANEH--YV--CQKGSFVAKTWTELLKHVRETHKE 215
Db 145 YKCHNE--CGKGFALPSRLIHRHTHTGKPYCSTECRKQ--FTMK--SDLIVHQ--THTA 197
OY 216 E--ILCEVCKTKRKRDYKQHKHTHAPRD--VCRSPREGGRATYTFNLQSHLSFHE 272
Db 198 EKPIKIDCKGKGLTVKSRILVHQHTHTGKPYVC--GEGCKGEPAKIRLMGHORT--HT 252
OY 273 ESRPFVCEHAGCGKTFAMKOSLYRNAVVDPPKKMKLKVKKSRKRS--LASHLSGYTPP 331
Db 253 GEKPYICNE--CGKGFTEKSHLVNHRHTHTGKPYVCEGCKGLTGKSLIAHQHTHTGE 310
OY 332 K-----RKQGGGLS-----CONGESP-NCVE-DK 354
Db 311 KPYICNECGKGFPMKSTLSIHQDTHTEGKPYCNCNCDK 348

RESULT 3

US-09-764-864-1045
: Sequence 1045, Application US/09764864
: Patent No. US20020132753A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT223
: CURRENT APPLICATION NUMBER: US/09/764,864
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1792
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 1045
: LENGTH: 457
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-764-864-1045

Query Match 19.1%; Score 383.5; DB 10; Length 457;
Best Local Similarity 30.8%; Pred. No. 1.8e-24;
Matches 123; Conservative 44; Mismatches 138; Indels 95; Gaps 19;

OY 41 ICSPDCCSANYSKAMKLDALHCKHTGERPVCDEGCGKAFIRDYHLSRHLTHTEGKPF 100
Db 45 VCT--ECGKAFKRSRLIYHQHTHTGKPHGCSI--CGKAFRSKGLTEHQRNHTGKEXP 100
OY 101 VCAATGCCDQKFNPKSKNLKKHFERKHEHQKQYIC----- 135
Db 101 EC--TECDKAFPMKSQLINAH--QKIHTEGKSYTCSCGCKGFIKRSRLIHQRVHTGKEXP 156
OY 136 -FEDCKKTFKKHQDLKIHQCOHTNEPLFKCTQEGSCGHFASPSKLRHAKANEH--YVC 191
Db 157 GCSLCKGAFKRSRLIYHQHTHTGKPYCTE--CDKAFPMKSQLINAHQKHTGKEXP 214
OY 192 QK--GCSFVAKTWTELLKHV--ETHKEPILCEVCKTKRKDYKQHKHTHAPRD--VCR 248
Db 215 KDCGKGFQK--GNLIVHQHTHTGKPYICNCGKGFQKGNLIIHRHTHTGKEXP 270
OY 249 PREGCGRTYTFVFNLSHLSFHEESRPFVCEHAGCGKTFAMKOSLYRNAVVDPPK--- 305
Db 271 --NECGKGFSGKTCGLISH--QRPHTGKTPVYCTE--CGKGSKSGSLIHRHTHTGKEXP 325
OY 306 -----KKMKLKVKKSRKRS-----LASHLS-----GYIPKRA----- 334
Db 326 CSDCKGKAFRDKSCLNHRHTHTGKPYGSDCKGKAFSHLSCLVYHKGMLHAREXCVGSVK 385
OY 335 -----QGGGLS-----CONGESPNCVEDKMLSTVAVLT 364
Db 386 LENPSESHTSHTRDILQKDSVNWVTLQMPSVAAQVSL 425

RESULT 4

US-09-976-165-10
: Sequence 10, Application US/09976165
: Patent No. US20020107383A1
: GENERAL INFORMATION:
: APPLICANT: FUJIMURA, TSUTOMU
: APPLICANT: MATANABE, TAKESHI
: APPLICANT: HORIE, MASATO
: TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
: FILE REFERENCE: Q-53599
: CURRENT APPLICATION NUMBER: US/09/976,165
: Prior application date: 2001-10-15
: Prior application number: 09/565,538
: Prior application date: 2000-05-05
: Prior application number: 09/055,699
: Prior application date: 1998-04-07
: Prior application number: 08/820,170
: Prior application date: 1997-03-19
: Prior application number: JP 63410/1996
: Prior application date: 1996-03-19
: Prior application number: JP 69163/1997
: Prior application date: 1997-03-05
: Number of seq id nos: 95
: Software: Patentln Ver. 2.1
: SEQ ID NO 10
: LENGTH: 711
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-976-165-10

Query Match 19.0%; Score 381; DB 10; Length 711;
Best Local Similarity 32.9%; Pred. No. 5e-24;

Matches 93; Conservative 43; Mismatches 119; Indels 28; Gaps 14;

OY 38 RRFICSPDCCSANYSKAMKLDALHCKHTGERPVCDEGCGKAFIRDYHLSRHLTHTEG 97
Db 445 KSYVCI--ECGQAFIQKANHILVHQHTHTGKPYOC--HNCGKSFISKSLDILHRIHTGE 500
OY 98 KPFVCAATGCCDQKFNPKSKNLKKHFERKHEHQKQYICSPEDCKKTFKKHQDLKIHQCOHT 157

```
Db 501 KPYEC--SDCGKFTTQKSHLNT--OKITGERHNHCS--ECGKAFNCKSTLSMHRIT 554
OY 158 NEPLFKCTOEGCGHNFPASKLRHAKAHG--YVOCGKSGFVAKTTELLKH-VRETH 213
Db 555 GEKPYKCS--CGKAFNKSQFKEHORITGKPYCTE--CGKAFNNGSNHFKHQTHTTR 611
OY 214 KEELICEVCKRTFKRKYLLKQNHKTAPENDYCRPRE--GGGRITTVFMLOSHILSFH 271
Db 612 ERPPVCTKCGKAFYOKSELTTHORTHNGE-----PYECLDGKSFSPKPOLKVH-ORIH 665
OY 272 EESRPVCEHAGCGKTFPAMKOSTLRHAYVHDPDK-KMKLYK 313
Db 666 TGERPYVCE--CGKAFNRSNPNKHQTHTTRDKSYKCSYK 706
```

```
RESULT 5
US-09-828-648-2
; Sequence 2, Application US/09828648
; Patent No. US2002015110A1
; GENERAL INFORMATION:
; APPLICANT: Ikezu, Tsuneya
; APPLICANT: Leisman, Gary
; APPLICANT: Carlson, Kimberly A.
; APPLICANT: Gendelman, Howard E.
; APPLICANT: University of Nebraska Medical Center
; TITLE OF INVENTION: Methods and Compositions for the
; TITLE OF INVENTION: Treatment of Human Immunodeficiency Virus Infection
; FILE REFERENCE: UMC 63156
; CURRENT APPLICATION NUMBER: US/09/828,648
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/246,331
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-828-648-2
```

```
Query Match 19.0%; Score 381; DB 10; Length 711;
Best Local Similarity 32.7%; Pred. No. 5e-24;
Matches 93; Conservative 43; Mismatches 119; Indels 28; Gaps 14;
OY 38 RRFICSPDGSANYSKAMKLDALHCKNHTGERPVCDEGCGKAFTRDYHLSHILHTG 97
Db 445 KSYVCI--ECGQAFITQKANHLYHQRSHTEKPYQC--HNGCKSFISKSQDLHHRIHGE 500
OY 98 KRPVCATGCDOKNTNSLNKHPERKHEKQKQYTCSEDECKTFKKHQQIKHQCOT 157
Db 501 KPYEC--SDCGKFTTQKSHLNT--OKITGERHNHCS--ECGKAFNCKSTLSMHRIT 554
OY 158 NEPLFKCTOEGCGHNFPASKLRHAKAHG--YVOCGKSGFVAKTTELLKH-VRETH 213
Db 555 GEKPYKCS--CGKAFNKSQFKEHORITGKPYCTE--CGKAFNNGSNHFKHQTHTTR 611
OY 214 KEELICEVCKRTFKRKYLLKQNHKTAPENDYCRPRE--GGGRITTVFMLOSHILSFH 271
Db 612 ERPPVCTKCGKAFYOKSELTTHORTHNGE-----PYECLDGKSFSPKPOLKVH-ORIH 665
OY 272 EESRPVCEHAGCGKTFPAMKOSTLRHAYVHDPDK-KMKLYK 313
Db 666 TGERPYVCE--CGKAFNRSNPNKHQTHTTRDKSYKCSYK 706
```

```
RESULT 6
US-09-864-761-33653
; Sequence 33653, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
```

```
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33653
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 284476.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BT4/4, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EST HUMAN HIT: AF150239.1, EVALUATE 1.00e-106
; OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALUATE 1.00e-106
US-09-864-761-33653
```

```
Query Match 18.9%; Score 379; DB 10; Length 525;
Best Local Similarity 32.7%; Pred. No. 5.1e-24;
Matches 101; Conservative 39; Mismatches 123; Indels 46; Gaps 13;
OY 47 CSANYSKAMKLDALHCKNHTGERPVCDEGCGKAFTRDYHLSHILHTGKRPVCAATG 106
Db 111 CGKNFSWMSDILHMQJHSGEKPRVCN--ECGKAFNTRQDLSMHRITHTGKRPFC--TQ 166
```

107 CDQKFNKSNLKKHFEKHNQOYICSPEDCKTEKKHLOOLKIHOCQHNPLFKCTQ 166
167 CGKAFNSRSLCRH--KTHISGEKPHEC--ROCGKAFTRNKLCHHOLHIGEKPRYKC-- 220
167 EGCGKHAFSPSKLRIKAKAHG--YVCGKGSFVAKTWTELLKHYR-ETKHEEILCEVC 222
221 NCCGKAFQFKSLITIHRIHTGKPYECE--CGKAFSGSSDLTKHRIHTGERPYEGSKC 279
223 KCTFKRDKYLOHKHTAPRDVCRPCGCGRTYTYVNLQSHLSFHEESPVECHA 282
280 GKAHSRSDLSKHKRITRKNH-GCPO--CGKPSIKRELTKH--RHHTTEKKRYCEEC 335
283 G-----CGKTFAMQSLTRHVVHDPDKKKMLKVKKS- 315
336 GKARHNCRRARHREHTGKPYQCRDCGCTGDKHCLTIHRIHTGKPYKLECGKAF 395
316 REKRSLSH 324
396 SGKSNLTNH 404

RESULT 7

US-09-764-864-948
Sequence 948, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 948
LENGTH: 338
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (335)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-948

Query Match 18.78; Score 375; DB 10; Length 338;
Best Local Similarity 31.58; Pred. No. 6,5e-24;
Matches 93; Conservative 49; Mismatches 125; Indels 28; Gaps 14;

38 RRTCSPPDCSANYSKAMLDANLCKHTGERPFVCDYEGCGKAFTRDYHLSRHILTHGE 97
12 KRMQCRM--CEKSTYSKYNLVTHILGNIGIKPHSCN--CSNLFQPSHILDTHTLHQGT 67
98 KPVYCATGCDQKNTNSLKKHFEKHNQOYICSPEDCKTEKKHLOOLKIHOCQHT 157
68 KPHKQV--GKATQTSILKRM--LHSEVKPYSCN--CGRGFAVPSSELKAHEVKHE 121
158 NEPLFKCTQEGCGKHAFSPSKLRIKAKAHG--YVCGKGSFVAKTWTELLKHYRE 211
122 SCGRHVCVE--CGIDFSTLTQLKRIHLASHOGPTLYOCLECDKSPHYRSQLONHMLK-- 176
212 THKEEILCEVCRTKFKKDKYLOHKHTAPRDVCRPCGCGRTYTYVNLQSHLSFHN 271
177 QNVRPFCTEGMEFQSILHKLKQSHLTHGVEF--KC--EYCGREFTLQANMKKHHML-IH 232
272 EESRPVCEHAGCGTFAMQSLTRHVVHDPDK--KMKLKVKKSREKRSLSH 325
233 TSVBRPQC-HI-CEKTFVQKQTLKTHMIVHSPVKPRKCKVCKSFNRMYNLGNH 285

RESULT 8

US-09-864-761-37065
Sequence 37065, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 37065
LENGTH: 551
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005261.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 2.3
OTHER INFORMATION: SWISSPROT HIT: P51542, EVALU0.00e+00
OTHER INFORMATION: EST HUMAN HIT: AF150239.1, EVALU0.100e-105
OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU0.100e-105
US-09-864-761-37065

Query Match 18.58; Score 372; DB 10; Length 551;
Best Local Similarity 30.28; Pred. No. 2.1e-23;
Matches 106; Conservative 45; Mismatches 124; Indels 76; Gaps 18;


```
OY 40 FICSPDGSANYSKAWKLDALCKHNGERPFVCDYEGCGKATRDYHLSRHILHTGKRP 99
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 127 FKCS--FCGKVFKNKHLIAGHEKIHSGVAPYEC--TECGKTYIKSTHLLQHHMHTGEP 182
OY 100 FVCAATGCDOKFNTKSNLKKHFRKHEHNOOKOYICSFEDCKKTFKKHQLIHOCQHTNE 159
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 183 YECME--CGKAFNRKSYLTOH-QRTHSG--EKPYKCN--BCGKAFTHRSNFVLIHNRHTGE 236
OY 160 PLFKCTOGS-----CGKHFASPSLKRHAHNS---YV 190
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 237 KSPVTECGGVFRFRKPGFLRHAYVHSGENPYECLECGKVFKRSTYLMHMOOHTGKPYE 296
OY 191 CGKCSFVAKTWELKLN-VRETHKEELICEVCRKTFKRKYDLKQHMKTAPERD-VCRC 248
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 297 CSB--CGKVFLESALLIHVYIHTGKRPFELECGKAFNRHRSYLRHORIHTGKRPVC-- 353
OY 249 PRECGRTYTVFVFLQSHIISF-----HESRPFVCEHAGCGKTFPAKOSLTRAVVHD 302
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 354 --SECGKAFI-----HCSTFLHKRAHTGKRPCKE--CGKAFSRKMDLIRHFSHT 402
OY 303 PDKRKMILKVKRSREKRSLSHLSGYIPPRKOGGLSLCQNGESP-NCVE 352
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 403 GEKPEYCEGCKA-----FTRMGSLTRHNR-----IHSGEKPYECVE 439

RESULT 9
US-09-864-761-37152
; Sequence 37152, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
```

```
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37152
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL049646.17
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EST_HUMAN HIT: AM953464.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P52736, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AM953464.1, EVALUE 0.00e+00

US-09-864-761-37152

Query Match 18.4%; Score 368.5; DB 10; Length 515;
Best Local Similarity 30.5%; Pred. No. 3,7e+23;
Matches 107; Conservative 45; Mismatches 118; Indels 81; Gaps 17;

OY 46 DCSANYSKAWKLDALCKHNGERPFVCDYEGCGKATRDYHLSRHILHTGKRPVCAAT 105
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 79 ECGLSFSKMTLLSHQNIHSGEKRPYGV--CGKGLSLKSLARHOKHSGEKRPYCKE- 135
OY 106 GCDOKFNTKSNLKKHFRKHEHNOOKOYICSFEDCKKTFKKHQLIHOCQHTNEPLFCT 165
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 136 -CGRFNFKRSLLIHT-ERTHSG-EKPYMCS--ECGRGFSQKSNLITHQTHSGEKRPYCR 190
OY 166 DECGCKHFPASPSLKRHAHNS-----EGVCCGK 194
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 191 E--CGKGFSSQKSAVVRHQRHLEKTYICSDGLGFSDRSNLISHQTHSGEKRPYAC-KE 247
OY 195 CSFVAKTWELKLVRETKEE--ILCEVCRKTFKRKYDLKQHMKTAPERD-VCRCRE 251
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 248 GCRCFRQRTLLVNHQR-THSEKRPYGVCGVGHFSSNTSLISHRTHGKRPYGVG-- 303
OY 252 CGGRTYTVFVFLQSHI--LSFHEESRPFVCEHAGCGKTFPAKOSLTRAVVHDPDK-- 305
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 304 -CGRG---FSLKSLHNRHQNHSGEKRPYCK--DCGRGFSQKSNLIRHQRTHSGEKPMV 356
OY 306 -----KMKLKVKK-----SREKRSLSHLSGYIPPRKRCQ 337
    11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db 357 CGECGRGFSQKSNLVAHQTHSGERPFVCDYEGCGKGFSDGGRGFSHQAGLIRKRRHSR 407

RESULT 10
US-09-864-761-42958
; Sequence 42958, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
```



```

: LOCATION: (290)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1065

Query Match
Best Local Similarity 29.3%; Score 366.5; DB 10; Length 547;
Matches 104; Conservative 47; Mismatches 116; Indels 81; Gaps 18;

QY 38 RRLTSPDGSANYSKAMKIDALCKHTGEPVCDYECGCAFLRDYHLSHLLITLITGE 97
DB 214 RYHICS--HCGKAFSQSISDLNRHOKHTGTDKPYKC-YE-CGKGFSSSHLSLQHOHTHTGE 269
QY 98 KPFVCAATG-----CDQKNTSKNKKKFERHENOOKQY 132
DB 270 RPYDCNCGSGSFGSSHLIXXHPRTGKPRHCKNEGSKSPCRSLHLQH-QRTHSQ-EKPY 327
QY 133 ICSFEDCKTKFKHQOLKTHQCOHTNEPLFKTOEGCGHFPASBPKLRHAKAHG--Y 189
DB 328 EC--EEGKSSFSRSHLAHORHTGTEKPEECNE--CGKGFSESDLLIKHRYHTGERPY 383
QY 190 VCGKCSFVAKTWTLLKHVR-ETRKEEILCEYCRKTEFKRYDKLQHNKTHAPERDVCRC 248
DB 384 KCDE--CGKNFSQNSDLYHRBRATGKPRHNCNCGENFSLHVLQHOHTHTGE----- 437
QY 249 PRE--CGGRITTYFNLSHLS--FHESRPVCE-----HAG-- 283
DB 438 PYECNAGGKS---FSRSHLLTHQKHTHGEKPYECNECWSFGERSDLIKHQRHTHTGEK 493
QY 284 -----CGKTFAMKOSLTRHVVHDPDKRKKMKLYKKSREKRS--LASH 324
DB 494 PYECVCGKGFQSSNLITHQVHTGKPYECTECKSPSSSALIKH 541

RESULT 13
US-09-989-920-212
: Sequence 212, Application US/09989920
: Patent No. US20020172957A1
: GENERAL INFORMATION:
: APPLICANT: Macina, Roberto
: APPLICANT: Recipon, Hervé
: APPLICANT: Chen, Sei-Yu
: APPLICANT: Sun, Yongming
: APPLICANT: Liu, Chenghua
: TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Pro
: FILE REFERENCE: DEX-0291
: CURRENT APPLICATION NUMBER: US/09/989,920
: PRIOR APPLICATION NUMBER: 2001-11-21
: PRIOR FILING DATE: 2000-11-22
: NUMBER OF SEQ ID NOS: 284
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 212
: LENGTH: 610
: TYPE: PRT
: ORGANISM: Homo sapien
US-09-989-920-212

Query Match
Best Local Similarity 18.2%; Score 366; DB 9; Length 610;
Matches 116; Conservative 52; Mismatches 136; Indels 124; Gaps 20;

QY 10 SVSSLTADAFIAGESSAPTPRPALPRRLTSPDGSANYSKAMKIDALCKHTGERP 69
DB 117 SYNSLILQHRIRHTGE-----RPYVCDV--CGKTFRNAGLKVHRHLHTGKPR 162
QY 70 FVCDYEGGKAFIDYHLSRIILTHGEKRPVCAATGCGOKNTSNLKKRPERHENOQ 129
DB 163 TKCDV--CGKATYSRSLKHNKGIHGEKPYKSY--CEKSPNYSALQDQ--RRIHTRE 216
QY 130 KQYICSEDDCKTKFKHQOLKTHQCOHTNEPLFKTOEGCGHFPASBPKLRHAKAHG 189
DB 217 KPFCC--DECGKAFRNNSGLKVKHRIHTGERPYKC--EEGKATYISLSLNLNHSVHGE 272

```

```

QY 190 -----VCORG-----CSFVAKT--WTELLKHVRETKEE--IL 218
DB 273 KPFKDECEKAFITYRTLTLNKKRVHIGERPYKDVCEKSFNTSLISOHRVHTREKPYE 332
QY 219 CEVCKTFKFRDYLOKMKTHAPER---DVC-----RCPR--EGC 253
DB 333 CDRCKEFVRNNSLKVAKRIHTGERPYECDVCGKATYISLSLNLNHSVHGE 392
QY 254 GRITYTYFNLSH-----ILSFHEE---SRPVCEHAGCK 286
DB 393 GKAFSSRTLLSHKRVHIGERPKVCEGKSPFSYSLISOHRHTGKRPVCDR--CGK 450
QY 287 TFAMKOSLTRHVVHDPDK-----KKMKLVKKSREKRSLSHLSGYIPRKRQOGISL 341
DB 451 AFRNSSGLTVKRIHTGKPYECDECGKATYISLS--SLNHSVH-----GKOPYN 500
QY 342 CGESGSPN 349
DB 501 CECKGSPN 508

RESULT 14
US-09-925-301-1357
: Sequence 1357, Application US/09925301
: Patent No. US200205208A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA106
: CURRENT APPLICATION NUMBER: US/09/925,301
: PRIOR FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05882
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1694
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1357
: LENGTH: 580
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (3)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (526)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1357

Query Match
Best Local Similarity 18.0%; Score 360.5; DB 10; Length 580;
Matches 92; Conservative 32; Mismatches 118; Indels 23; Gaps 13;

QY 46 DCSANYSKAMKIDALCKHTGEPVCDYECGKAFIRDYHLSRIILTHGEKRPVCAAT 105
DB 247 ECGKGFRTSATLNTHTGTHGNDNYEC--KEGKATFRSCQILOHRIKHTGKPRYKC--K 302
QY 106 GCDOKFNTKSNLKHNFERRKHENOOKOYICSFEDCKTKFKHQOLKTHQCOHTNEPLFKCT 165
DB 303 DCGRAFTVSSLSQDM--KIHVEKEPYEC--KECGIAFTRSSQLEHNLKTHAKDPFEC- 357
QY 166 OEGCGKHPASPSKLRHAKAHG---YVCGKCSFVAKTWTLLKHVHTKEE--IICE 220
DB 358 -KICGKSPRNSCLSDHFRHTHTGKPYKC-KDCGKAFYTONSDLTTHAR--THSGEPYCK 414
QY 221 VCRKTEKRDYLOKMKTHAPERDVCPRGSGRTTYTVFNLSHLSLFSHESRPVCE 280
DB 415 ECGKAFARNSRLSEHTRTHTGKPRF-ECYK--CGKAFATSSNLSDH--LRIHTGKPRFEC- 469
QY 281 HAGCGKTFAMKOSLTRHVVHDPDK 305
DB 470 -LECGKATFTHSSSLNHNKTHSAK 493

```



```
Db      61 LCKHTGERPPVCDYEGCGKAFIRDYHLRSRHLTHNGEKPRFVCAATGCDQCKFNYSNLKKH 120
OY      121 FERKHENOQOYICSFEDCKKTKRKHQQLKHQCOHTNBPFLKCTQEGCGKHFAFPSKXK 180
Db      121 FERKHENOQOYICSFEDCKKTKRKHQQLKHQCOHTNBPFLKCTQEGCGKHFAFPSKXK 180
OY      181 RHAKAHHEGYVOCOKCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRDYLOKOMKTHA 240
Db      181 RHAKAHHEGYVOCOKCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRDYLOKOMKTHA 240
OY      241 PERDYCPCREGCGRTYTVFNLOSHLSFHEHSRPFVCEHAGCGKTFAMQSLTRHAVV 300
Db      241 PERDYCPCREGCGRTYTVFNLOSHLSFHEHSRPFVCEHAGCGKTFAMQSLTRHAVV 300
OY      301 HDPOKXKKKLLVKKSRKRSLSLASHLSGYIPPKRKQOGSLCQNGESPNCVEDKMLSTVA 360
Db      301 HDPOKXKKKLLVKKSRKRSLSLASHLSGYIPPKRKQOGSLCQNGESPNCVEDKMLSTVA 360
OY      361 VLTIG 365
Db      361 VLTIG 365

RESULT 2
US-09-831-426c-2
; Sequence 2, Application US/09831426C
; GENERAL INFORMATION:
; APPLICANT: Hoechst Marion Roussel
; APPLICANT: Bordon-Pailler, F.
; APPLICANT: Rocher, C.
; TITLE OF INVENTION: Human htf111a gene and coded htf111a protein
; FILE REFERENCE: 146.1364
; CURRENT APPLICATION NUMBER: US/09/831,426C
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 2
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Human
US-09-831-426c-2

Query Match
Best Local Similarity 100.0%; Score 2008; DB 22; Length 365;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 MDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPFDCSANYSKAMKLDAAH 60
Db      1 MDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPFDCSANYSKAMKLDAAH 60
OY      61 LCKHTGERPPVCDYEGCGKAFIRDYHLRSRHLTHNGEKPRFVCAATGCDQCKFNYSNLKKH 120
Db      61 LCKHTGERPPVCDYEGCGKAFIRDYHLRSRHLTHNGEKPRFVCAATGCDQCKFNYSNLKKH 120
OY      121 FERKHENOQOYICSFEDCKKTKRKHQQLKHQCOHTNBPFLKCTQEGCGKHFAFPSKXK 180
Db      121 FERKHENOQOYICSFEDCKKTKRKHQQLKHQCOHTNBPFLKCTQEGCGKHFAFPSKXK 180
OY      181 RHAKAHHEGYVOCOKCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRDYLOKOMKTHA 240
Db      181 RHAKAHHEGYVOCOKCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRDYLOKOMKTHA 240
OY      241 PERDYCPCREGCGRTYTVFNLOSHLSFHEHSRPFVCEHAGCGKTFAMQSLTRHAVV 300
Db      241 PERDYCPCREGCGRTYTVFNLOSHLSFHEHSRPFVCEHAGCGKTFAMQSLTRHAVV 300
OY      301 HDPOKXKKKLLVKKSRKRSLSLASHLSGYIPPKRKQOGSLCQNGESPNCVEDKMLSTVA 360
Db      301 HDPOKXKKKLLVKKSRKRSLSLASHLSGYIPPKRKQOGSLCQNGESPNCVEDKMLSTVA 360
OY      361 VLTIG 365
Db      361 VLTIG 365
```

```
RESULT 3
PCT-US00-26524B-5945
; Sequence 5945, Application PC/TUS0026524B
; GENERAL INFORMATION:
; APPLICANT: Birse et. al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005PCT
; CURRENT APPLICATION NUMBER: PCT/US00/26524B
; CURRENT FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5945
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (3)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
PCT-US00-26524B-5945

Query Match
Best Local Similarity 99.5%; Score 2000; DB 1; Length 409;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY      1 MDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPFDCSANYSKAMKLDAAH 60
Db      45 LDPRAVVAESVSLTIADAFIAGESSAPTPPPALPRRFICSPFDCSANYSKAMKLDAAH 104
OY      61 LCKHTGERPPVCDYEGCGKAFIRDYHLRSRHLTHNGEKPRFVCAATGCDQCKFNYSNLKKH 120
Db      105 LCKHTGERPPVCDYEGCGKAFIRDYHLRSRHLTHNGEKPRFVCAATGCDQCKFNYSNLKKH 164
OY      121 FERKHENOQOYICSFEDCKKTKRKHQQLKHQCOHTNBPFLKCTQEGCGKHFAFPSKXK 180
Db      165 FERKHENOQOYICSFEDCKKTKRKHQQLKHQCOHTNBPFLKCTQEGCGKHFAFPSKXK 224
OY      181 RHAKAHHEGYVOCOKCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRDYLOKOMKTHA 240
Db      225 RHAKAHHEGYVOCOKCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRDYLOKOMKTHA 284
OY      241 PERDYCPCREGCGRTYTVFNLOSHLSFHEHSRPFVCEHAGCGKTFAMQSLTRHAVV 300
Db      285 PERDYCPCREGCGRTYTVFNLOSHLSFHEHSRPFVCEHAGCGKTFAMQSLTRHAVV 344
OY      301 HDPOKXKKKLLVKKSRKRSLSLASHLSGYIPPKRKQOGSLCQNGESPNCVEDKMLSTVA 360
Db      345 HDPOKXKKKLLVKKSRKRSLSLASHLSGYIPPKRKQOGSLCQNGESPNCVEDKMLSTVA 404
OY      361 VLTIG 365
Db      405 VLTIG 409

RESULT 4
US-10-106-698-5955
; Sequence 5955, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
```

```

: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: Patentln Ver. 3.0
: SEQ ID NO 5955
: LENGTH: 409
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: MISC_FEATURE
: LOCATION: (3)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5955

Query Match
Best Local Similarity 99.5%; Score 2000; DB 25; Length 409;
Matches 363; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDPAAVAESVSLTTADAFIAGESSAPTPRRPALPRPFTCSFPDCCSANTSKAMKLDAN 60
Db :|||||
45 LDPAAVAESVSLTTADAFIAGESSAPTPRRPALPRPFTCSFPDCCSANTSKAMKLDAN 104
QY 61 LCKHTGERPVCYDEGCGNAFIRDYHLNRHILTHTGERPVCATGCDQKFKNTSKNKKH 120
Db :|||||
105 LCKHTGERPVCYDEGCGNAFIRDYHLNRHILTHTGERPVCATGCDQKFKNTSKNKKH 164
QY 121 PERKHEMQQOYICSEFDCKKTPKKNQQLKIQOCOTNBPYFKCTQEGCGKHPASPSK 180
Db :|||||
165 PERKHEMQQOYICSEFDCKKTPKKNQQLKIQOCOTNBPYFKCTQEGCGKHPASPSK 224
QY 181 RHAAHNGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNQHMKT 240
Db :|||||
225 RHAAHNGYVCGKGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNQHMKT 284
QY 241 PERVYCGRPREGGRTTYVFNLSHISFHEESRPVCEHAGCGKTFAMKOSLTRHAY 300
Db :|||||
285 PERVYCGRPREGGRTTYVFNLSHISFHEESRPVCEHAGCGKTFAMKOSLTRHAY 344
QY 301 HDPDKKMKLVKRSREKSLASHLSGYTPRKKGQGLSLCQNGESPNCVEDMLSTVA 360
Db :|||||
345 HDPDKKMKLVKRSREKSLASHLSGYTPRKKGQGLSLCQNGESPNCVEDMLSTVA 404
QY 361 VLTIG 365
Db :|||||
405 VLTIG 409

RESULT 5
PCT-US01-18569-2978
: Sequence 2978, Application PC/TUS0118569
: GENERAL INFORMATION:
: APPLICANT: Human Genome Sciences, Inc.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA133PCT
: CURRENT APPLICATION NUMBER: PCT/US01/18569
: CURRENT FILING DATE: 2001-06-07
: PRIOR APPLICATION NUMBER: 60/209,467
: PRIOR FILING DATE: 2000-06-07
: NUMBER OF SEQ ID NOS: 4360
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 2978
: LENGTH: 184
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (26)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (36)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (123)
```

```

: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-18569-2978

Query Match
Best Local Similarity 40.9%; Score 821; DB 1; Length 184;
Matches 155; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 193 KGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNQHMKTAPEDVCRCPREG 252
Db :|||:|||||
12 KRMSLCGRTWADLXHVREPIKREXLCGVCRKTFAMKDYLNQPMKTHAPEDVCRCPREG 71
QY 253 CGRTTYVFNLSHISFHEESRPVCEHAGCGKTFAMKOSLTRHAYVHDPDKKMKLV 312
Db :|||||
72 CGRTTYVFNLSHISFHEESRPVCEHAGCGKTFAMKOSLTRHAYVHDPDKKMKLV 131
QY 313 KKSREKSLASHLSGYTPRKKGQGLSLCQNGESPNCVEDMLSTVAVLTIG 365
Db :|||||
132 KKSREKSLASHLSGYTPRKKGQGLSLCQNGESPNCVEDMLSTVAVLTIG 184

RESULT 6
US-10-264-049-2978
: Sequence 2978, Application US/10264049
: GENERAL INFORMATION:
: APPLICANT: Birce et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PA133P1
: CURRENT APPLICATION NUMBER: US/10/264,049
: CURRENT FILING DATE: 2002-10-04
: PRIOR APPLICATION NUMBER: PCT/US01/18569
: PRIOR FILING DATE: 2001-06-07
: PRIOR APPLICATION NUMBER: US 60/209,467
: PRIOR FILING DATE: 2000-06-07
: NUMBER OF SEQ ID NOS: 4360
: SOFTWARE: Patentln Ver. 3.1
: SEQ ID NO 2978
: LENGTH: 184
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: MISC_FEATURE
: LOCATION: (26)
: OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
: NAME/KEY: MISC_FEATURE
: LOCATION: (36)
: OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
: NAME/KEY: MISC_FEATURE
: LOCATION: (123)
: OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2978

Query Match
Best Local Similarity 89.6%; Score 821; DB 26; Length 184;
Matches 155; Conservative 1; Mismatches 17; Indels 0; Gaps 0;

QY 193 KGSFVAKTWTLLKHVRETHKEEILCEVCRKTFKRDYLNQHMKTAPEDVCRCPREG 252
Db :|||:|||||
12 KRMSLCGRTWADLXHVREPIKREXLCGVCRKTFAMKDYLNQPMKTHAPEDVCRCPREG 71
QY 253 CGRTTYVFNLSHISFHEESRPVCEHAGCGKTFAMKOSLTRHAYVHDPDKKMKLV 312
Db :|||||
72 CGRTTYVFNLSHISFHEESRPVCEHAGCGKTFAMKOSLTRHAYVHDPDKKMKLV 131
QY 313 KKSREKSLASHLSGYTPRKKGQGLSLCQNGESPNCVEDMLSTVAVLTIG 365
Db :|||||
132 KKSREKSLASHLSGYTPRKKGQGLSLCQNGESPNCVEDMLSTVAVLTIG 184

RESULT 7
US-09-760-466-1024
: Sequence 1024, Application US/09760466
```

```

: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT259
: CURRENT APPLICATION NUMBER: US/09/760,466
: CURRENT FILING DATE: 2001-01-16
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1813
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 1024
: LENGTH: 106
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-760-466-1024

Query Match
Best Local Similarity 27.7%; Score 556; DB 21; Length 106;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 VFNLQSHILSFHESRPFVCEHAGCGKTFAMQSLTRHAYVHDPKKMKLKVKKSREKR 319
DB 1 VFNLQSHILSFHESRPFVCEHAGCGKTFAMQSLTRHAYVHDPKKMKLKVKKSREKR 60
DB 61 SLASHLSGYIPPRKQGGSLCQNGESPNCVEDKMLSTVAVLTG 106

RESULT 8
US-09-760-494-255
: Sequence 255, Application US/09760494
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC018
: CURRENT APPLICATION NUMBER: US/09/760,494
: CURRENT FILING DATE: 2001-01-16
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 258
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 255
: LENGTH: 106
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-760-494-255

Query Match
Best Local Similarity 27.7%; Score 556; DB 21; Length 106;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 VFNLQSHILSFHESRPFVCEHAGCGKTFAMQSLTRHAYVHDPKKMKLKVKKSREKR 319
DB 1 VFNLQSHILSFHESRPFVCEHAGCGKTFAMQSLTRHAYVHDPKKMKLKVKKSREKR 60
QY 320 SLASHLSGYIPPRKQGGSLCQNGESPNCVEDKMLSTVAVLTG 365
DB 61 SLASHLSGYIPPRKQGGSLCQNGESPNCVEDKMLSTVAVLTG 106

RESULT 9
US-10-212-083-1024
: Sequence 1024, Application US/10212083
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT259C1N
: CURRENT APPLICATION NUMBER: US/10/212,083
: CURRENT FILING DATE: 2002-08-06
: Prior application data removed - See file wrapper or PALM.
: Prior application number: 09/760,466
: Prior application number: 2001-01-16
: Prior application number: 60/179,065
: Prior application number: 2000-01-31
: Prior application number: 60/180,628
```

```

: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: 60/214,886
: PRIOR FILING DATE: 2000-06-28
: PRIOR APPLICATION NUMBER: 60/217,487
: PRIOR FILING DATE: 2000-07-11
: PRIOR APPLICATION NUMBER: 60/225,758
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/220,963
: PRIOR FILING DATE: 2000-07-26
: PRIOR APPLICATION NUMBER: 60/217,496
: PRIOR FILING DATE: 2000-07-11
: PRIOR APPLICATION NUMBER: 60/225,447
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/218,290
: Remaining prior application data removed - See file wrapper or PALM.
: NUMBER OF SEQ ID NOS: 1813
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 1024
: LENGTH: 106
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-212-083-1024

Query Match
Best Local Similarity 27.7%; Score 556; DB 26; Length 106;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 260 VFNLQSHILSFHESRPFVCEHAGCGKTFAMQSLTRHAYVHDPKKMKLKVKKSREKR 319
DB 1 VFNLQSHILSFHESRPFVCEHAGCGKTFAMQSLTRHAYVHDPKKMKLKVKKSREKR 60
QY 320 SLASHLSGYIPPRKQGGSLCQNGESPNCVEDKMLSTVAVLTG 365
DB 61 SLASHLSGYIPPRKQGGSLCQNGESPNCVEDKMLSTVAVLTG 106

RESULT 10
US-10-223-026-255
: Sequence 255, Application US/10223026
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC018C1N
: CURRENT APPLICATION NUMBER: US/10/223,026
: CURRENT FILING DATE: 2002-08-19
: PRIOR APPLICATION NUMBER: 09/760,494
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: 60/179,065
: PRIOR FILING DATE: 2000-01-31
: PRIOR APPLICATION NUMBER: 60/180,628
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: 60/214,886
: PRIOR FILING DATE: 2000-06-28
: PRIOR APPLICATION NUMBER: 60/217,487
: PRIOR FILING DATE: 2000-07-11
: PRIOR APPLICATION NUMBER: 60/225,758
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/220,963
: PRIOR FILING DATE: 2000-07-26
: PRIOR APPLICATION NUMBER: 60/217,496
: PRIOR FILING DATE: 2000-07-11
: PRIOR APPLICATION NUMBER: 60/225,447
: PRIOR FILING DATE: 2000-08-14
: PRIOR APPLICATION NUMBER: 60/218,290
: Remaining prior application data removed - See file wrapper or PALM.
: NUMBER OF SEQ ID NOS: 258
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 255
: LENGTH: 106
: TYPE: PRT
: ORGANISM: Homo sapiens
```


;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: 09/496,914
;; PRIOR FILING DATE: 2000-02-03
;; NUMBER OF SEQ ID NOS: 3960
;; SOFTWARE: Custom
;; SEQ ID NO 1961
;; LENGTH: 869
;; TYPE: PRT
;; ORGANISM: Homo sapiens
PCT-US01-04098A-1961

Query Match 21.0%; Score 421; DB 1; Length 869;
Best Local Similarity 32.0%; Pred. No. 8,6e-27;
Matches 116; Conservative 41; Mismatches 116; Indels 89; Gaps 19;

OY 3 PPAVASVSLLTIADAFIAGESSAPTPPPRA-LPRRFICFPDCCSANYSKAMKLDHL 61
DB 125 PPVPAPEPRLRELVOG-----RRAGAERPTCN--ECGSEFSQMSKLRHQ 169
OY 62 CKHTGERPVCDEEGCGKAFIRYHLNRILTHTEGKPPVC----- 102
DB 170 RIHTGERPNTC--SECKSFTQSSHLVQHQRTHTEGKPPVCDCCKCFSSWSNLVQHQR 227
OY 103 -----AATGCDOKFNTKSLKKHFERKHENQKQYICSEFDCCKTFKKHQDLKHOQ 155
DB 228 HTGERPVCDEEGCGKAFIRYHLNRILTHTEGKPPVCDCCKCFSSWSNLVQHQR 283
OY 156 HTNEPRLKCTQEGCGKAFIRYHLNRILTHTEGKPPVCDCCKCFSSWSNLVQHQR 203
DB 284 HTGERPVCDEEGCGKAFIRYHLNRILTHTEGKPPVCDCCKCFSSWSNLVQHQR 341
OY 204 -----ELKKHRETKEE--ILCEVCRTFKKDYLNKQMTNAPER 243
DB 342 GEKPYECLEGGKSHGSHSTLIKQNR-THLRDEPFCPCVCGKFTLSATLLNKHQTHR 400
OY 244 DVCGRPEGCGRTTYTFNLSHLSFHEESRPVCEHAAGCKTAMKQSLTRIAVYHDP 303
DB 401 PY-KCPE--CGKRFSGNHLNKHQTHRDEPFCPCVCGKFTLSATLLNKHQTHR 454

OY 304 DK 305
DB 455 EK 456

RESULT 14
PCT-US01-04098A-1676
Sequence 1676, Application PC/TUS0104098A
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 21272-029
CURRENT APPLICATION NUMBER: PCT/US01/04098A
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: Not Yet Assigned
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom

;; SEQ ID NO 1676
;; LENGTH: 803
;; TYPE: PRT
;; ORGANISM: Homo sapiens
PCT-US01-04098A-1676

Query Match 20.8%; Score 417.5; DB 1; Length 803;
Best Local Similarity 29.1%; Pred. No. 1,6e-26;
Matches 111; Conservative 47; Mismatches 131; Indels 93; Gaps 14;

OY 18 DAFAAGSSAPTP-----PRPALPRR-----FTCSFPDCCSANYSKAMKLD 58
DB 230 DLLAEPAEPAPAPAPPEEAEGPAALGPRGLSGPGVLYLCEACCGGTFAKKHQLK 289
OY 59 AHLCKNT---GERPVCDEEGCGKAFIRYHLNRILTHTEGKPPVCDCCKCFSSWSN 115
DB 290 VHLTHSSQGRFRFCPLGCGMTFTTYSKLRHLSHDKLRPFGCAECCGCSFTTY 349
OY 116 NLKHF-----ERKHENQKQYICSEFDCCKTFKKHQ 148
DB 350 NLKHF-----ERKHENQKQYICSEFDCCKTFKKHQ 409
OY 149 LKHQOCHTNE-PLFCTQEGCGKHFASPSKLRHAKAHG--YVCG-KGCSVAATWT 203
DB 410 LFSNRHAFRQELFSCFPGCSKQYDKACRLKILHSHTRPRLCPDCCGNNFTS 469
OY 204 ELKKHRETKEEILCEVCRTFKKDYLNKQMTNAPERDVCRCPRGCGRTTYTFN 263
DB 470 KLRH-----KRR-----HDDRPF-MCPVGGCGKSFTRAEHL 501
OY 264 QSHLSFHEESRPVCEHAAGCKTAMKQSLTRIAVYH--DDPKMKMLKV---KKSRE 317
DB 502 KGHSTT-HLGTTRPVCVAGCCARFASRSLYHSHKHLQDVTMKSNCPISSCNKLFTS 560
OY 318 KRLSLASHSGTTPPKRQOG 339
DB 561 KSMKTHM---VNRHKGODL 578

RESULT 15
US-10-029-386-32185
Sequence 32185, Application US/10029386
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: AROMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
PRIOR FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32185
LENGTH: 803
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO Z99130.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
OTHER INFORMATION: SWISSPROT HIT: P98169, EVALU 0.00e+00
US-10-029-386-32185

Query Match 20.8%; Score 417.5; DB 24; Length 803;
Best Local Similarity 29.1%; Pred. No. 1,6e-26;
Matches 111; Conservative 47; Mismatches 131; Indels 93; Gaps 14;

```
OY 18 DAFIAGSSAPTP-----PRPALPR-----FICSPDCSANYSKAMKLD 58
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 DILAEPAPAPABAPBEEBAEGPAAALGPRGLSGPGVYLTCPEAOGOTFAKKHOLK 289
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 59 AHLCKNT---GERPVCDEYEGCGAFIRDYHLNRHILHTTGKPFVCAATGCDOKFNTKS 115
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 290 VHLTHSSSGORBPFGKPLGCGWTFFTSYKLRHLSHDKLRPFPCPAEGGKSFITYV 349
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 116 NLKKHF-----ERKHENOKOYICSFEDCKKTFKKHQ 148
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 350 NLKAMKCHGEENSFKCEVESEFPFOAKLSAHORSHFEPEKPYOCASFSGCKTFITYSA 409
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 149 LKHOCQHTNE-PLFKCTOEGCGNHFASPSKLRHAKAHG---YVCO-KGCSFYAKTWT 203
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 410 LFSHRAHFRQELFSCSPGCSKQYDKACRLKILRSHGTGERPFLDDEGCGNFTSMS 469
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 204 ELKKVRETHKEELICVCKRTFKRKDYLKQHMKTAPRDYCRCPREGCGRTYTTVNL 263
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 470 KILRH-----KHK-----HDDRRF-MCPVEGCGKSFTRAHEL 501
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 264 QSHILSFHEESRPVCEHAGCGKTFAMKOSLTRHAYVH--DDPKRMKLV---KKSRE 317
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 502 KGHST-HLGTKPFVCPVAGCAAFSARSSLYIHSKKHLQDVDTWMSRCPISSCNKLFTS 560
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 318 KRSLSHLSGYTPPKRKOGGL 339
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 561 KHSMTM---VKRHKYGDDL 578
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: February 10, 2003, 17:55.41
Job time : 304 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 17:02:49 : Search time 41 Seconds
(without alignments)
855.831 Million cell updates/sec

Title: US-09-831-426c-2
Perfect score: 2008
Sequence: 1 MDPFVAESVSLTIADAF.....ESPNCVEDKMLSTVAVLTLG 365

Scoring table: ELOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR-73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1992	99.2	363	2	DNA/RNA-binding pr
2	1814.5	90.4	338	2	transcription fact
3	1051.5	52.4	344	1	transcription fact
4	1047	52.1	335	2	transcription fact
5	1031	51.3	339	2	transcription fact
6	1022	50.9	339	2	transcription fact
7	556.5	27.7	365	2	transcription fact
8	553.5	27.6	365	2	transcription fact
9	496	24.7	365	2	transcription fact
10	409	20.4	336	2	transcription fact
11	406	20.2	693	2	transcription fact
12	404	20.1	632	2	transcription fact
13	398.5	19.8	626	2	transcription fact
14	397.5	19.8	591	2	transcription fact
15	394.5	19.6	803	2	transcription fact
16	392.5	19.5	337	2	transcription fact
17	391	19.5	651	2	transcription fact
18	386.5	19.2	457	2	transcription fact
19	386.5	19.2	615	2	transcription fact
20	385	19.0	600	2	transcription fact
21	382.5	19.0	280	2	transcription fact
22	381.5	19.0	485	2	transcription fact
23	380.5	18.9	595	2	transcription fact
24	377.5	18.8	595	2	transcription fact
25	375.5	18.7	496	2	transcription fact
26	375	18.7	710	2	transcription fact
27	374.5	18.7	515	2	transcription fact
28	374.5	18.7	1350	2	transcription fact
29	374	18.6	701	2	transcription fact

30	371	18.5	728	2	A48830	probable transcrip
31	371	18.5	1191	2	S35305	zinc finger protei
32	368.5	18.4	654	2	A57785	finger protein ZNF
33	368.5	18.4	673	2	S35335	transcription fact
34	368	18.3	435	2	S00833	finger protein (cl
35	367.5	18.3	576	2	A48157	renal transcrip
36	367	18.3	675	2	S51037	zinc-finger protei
37	366.5	18.3	428	2	A32891	finger protein 1,
38	365.5	18.2	399	2	S47071	finger protein H2F
39	365	18.2	580	2	A37107	spermatogenesis pr
40	364.5	18.2	589	2	I38598	zinc finger protei
41	363.5	18.1	393	2	JN0533	finger protein pml
42	363.5	18.1	469	2	I38600	zinc finger protei
43	363	18.1	594	2	T12488	hypothetical prote
44	362.5	18.1	553	2	S22954	finger protein zfp
45	362.5	18.1	688	2	A56360	zinc finger protei

ALIGNMENTS

RESULT 1

DNA/RNA-binding protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I38937

R:Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.E.; Becker, K.G.
Gene 159, 215-218, 1995
A:Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIIA
A:Reference number: I38937, MIMID:95347600, PMID:762052
A:Accession: I38937
A>Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-363 <RES>
A:Cross-references: EMBL:020272; NID:6644870; PIDN:AAA75623.1; PID:6644871
C:Superfamily: transcription factor IIIA

Query Match 99.2% Score 1992: DB 2: Length 363:
Best local similarity 99.7% Pred No. 1.2e-143:
Matches 362, Conservative 0, Mismatches 1, Indels 0, Gaps 0:

QY	3	PPAVVAESVSLTIADAFIAGESSAPTPRALPRFICSEPPDCSANYSKAMLDALIC	62
DB	1	PPAVVAESVSLTIADAFIAGESSAPTPRALPRFICSEPPDCSANYSKAMLDALIC	60
QY	63	KHTGERPVCDEGCGAFIRDYHLNRHILTHGKPFVCAATGCDOKFNKSLKXHE	122
DB	61	KHTGERPVCDEGCGAFIRDYHLNRHILTHGKPFVCAATGCDOKFNKSLKXHE	120
QY	123	RKHENQOXYICSEFEDCKTFKKHQDLKHOCNTNEPLFKCTOEGCGKHEASPSKLRH	182
DB	121	RKHENQOXYICSEFEDCKTFKKHQDLKHOCNTNEPLFKCTOEGCGKHEASPSKLRH	180
QY	183	AKAHGVCYCGKGSFVAKTWTETLKHVRETHKEILCEVCRKTFRKXYLKHQHKTAPE	242
DB	181	AKAHGVCYCGKGSFVAKTWTETLKHVRETHKEILCEVCRKTFRKXYLKHQHKTAPE	240
QY	243	RDVRCRPECGRRTYTFVNLQSHILSFHDESRPFVCEHAGCGKTFPAMKOSLTRHAYVD	302
DB	241	RDVRCRPECGRRTYTFVNLQSHILSFHDESRPFVCEHAGCGKTFPAMKOSLTRHAYVD	300
QY	303	PDKKMKLTKVKKSEKRSLSHSGYIPPRKGGGLSLCNGCEPNCVEDEKMLSTVAVL	362
DB	301	PDKKMKLTKVKKSEKRSLSHSGYIPPRKGGGLSLCNGCEPNCVEDEKMLSTVAVL	360
QY	363	TLG 365	
DB	361	TLG 363	
RESULT 2			
G01496			

transcription factor IIIA - human (fragment)
N:Alternate names: TFIIIA
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C:Accession: G01496
R:Becker, K.G.
submitted to the EMBL Data Library, August 1994
A:Reference number: G07520
A:Accession: G01496
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-338 <DBC>
A:Cross-references: EMBL:014134; NID:9551534; PIDN:AAA21873.1; PID:9551535
C:Genetics:
A:Gene: GDB:GTF3A; TFIIIA
A:Cross-references: GDB:434744; OMIM:600860
A:Map position: 13q12.3-13q13.1
C:Superfamily: transcription factor IIIA

Query Match 90.4%; Score 1814.5; DB 2; Length 338;
Best Local Similarity 92.3%; Pred. No. 3,2e-130;
Matches 335; Conservative 0; Mismatches 3; Indels 25; Gaps 1;

QY 3 PPAVAESVSLTADAFIAGESSAPTPRPALPRFICSPDGSANYSKAMKLDALHC 62
DB 1 PPAVAESVSLTADAFIAGESSAPTPRPALPRFICSPDGSANYSKAMKLDALHC 60
QY 63 KHTGERPVCDYECGCAFRDYHLSRHLHTGKRPVCAATGCDOKFNTKSLKHPF 122
DB 61 KHTGERPVCDYECGCAFRDYHLSRHLHTGKRPVCAATGCDOKFNTKSLKHPF 120
QY 123 RKNHNOQKQYICSFEDCKTKFKKHQDLKHOCQHTNEPLFKCTGECGCKHSPSKLRH 182
DB 121 RKNHNOQKQYICSFEDCKTKFKKHQDLKHOCQHTNEPLFK----- 161
QY 183 AKAHBGVYCKGCSFVAKTTELKKNVRETHKEEITCEYCKRTFKRKQYTLQNHKTAPE 242
DB 162 -----LCCQGGCSFVAKTTELKKNVRETHKEEITCEYCKRTFKRKQYTLQNHKTAPE 215
QY 243 RDVSCRPRGCGRTYTVFNLQSHLSFHEESRPFVCSNACGCKTFAKQSLTRHAAVND 302
DB 216 RDVSCRPRGCGRTYTVFNLQSHLSFHEESRPFVCSNACGCKTFAKQSLTRHAAVND 275
QY 303 PDKKKMLTKVKKSEKRSLSHLSGTYPRKQGGSLSCQNGESPNCVCKMLSTVAVL 362
DB 216 PDKKKMLTKVKKSEKRSLSHLSGTYPRKQGGSLSCQNGESPNCVCKMLSTVAVL 335
QY 363 TLG 365
DB 336 TLG 338

RESULT 3

transcription factor IIIA - African clawed frog

N:Alternate names: factor A; TFIIIA

C:Species: Xenopus laevis (African clawed frog)

C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 21-Jul-2000

C:Accession: A90857; A24961; S21776; A03529; A91007; S40784; S40785

R:Ginsberg, A.M.; King, B.O.; Roeder, R.G.

Cell 39, 479-489, 1984

A:Title: Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone

A:Reference number: A90857; MUID:85074456; PMID:6210149

A:Accession: A90857

A:Molecule type: mRNA

A:Residues: 1-344 <GIN>

A:Cross-references: GB:K02938; NID:9214818; PIDN:AAA49967.1; PID:9214819

R:Miller, J.; Macdonald, A.D.; Klug, A.

EMBO J. 4, 1609-1614, 1985

A:Title: Repetitive zinc-binding domains in the protein transcription factor IIIA from X

A:Reference number: A91007; MUID:85284956; PMID:4040853

A:Contents: annotation

R:Smith, J.F.; Hawkins, J.; Leonard, R.E.; Hanas, J.S.

Nucleic Acids Res. 19, 6871-6876, 1991

A:Title: Structural elements in the N-terminal half of transcription factor IIIA regu

A:Reference number: S40785; MUID:92107675; PMID:1762917

A:Contents: annotation; site-directed mutagenesis

R:Tso, J.Y.; Van Den Berg, D.J.; Korn, L.J.

Nucleic Acids Res. 14, 2187-2200, 1986

A:Title: Structure of the gene for Xenopus transcription factor TFIIIA.

A:Reference number: A24961; MUID:86167622; PMID:3754326

A:Accession: A24961

A:Molecule type: DNA

A:Residues: 1-296; C, 298-312, T, 314-333, D, 335-342, L, 344 <TSO>

A:Cross-references: EMBL:X03681; NID:965111; PIDN:CB81745.1; PID:95679716

R:Lillo, X.; Clemens, K.R.; Tennant, L.; Wright, P.E.; Gottesfeld, J.M.

J. Mol. Biol. 223, 857-871, 1992

A:Title: Specific interaction of the first three zinc fingers of TFIIIA with the inte

A:Reference number: S21776; MUID:92167265; PMID:1538401

A:Accession: S21776

A:Molecule type: mRNA

A:Residues: 1-101 <LIA>

R:del Rio, S.; Setzer, D.R.

Nucleic Acids Res. 19, 6197-6203, 1991

A:Title: High yield purification of active transcription factor IIIA expressed in E.c

A:Reference number: S40784; MUID:92066468; PMID:1956778

A:Contents: annotation

A:Note: an engineered sequence was used to obtain high-yield expression in an E. coli

these genes by RNA polymerase III. It also binds the transcribed 5S RNAs.

C:Comment: Most of the Cys and His residues are ligands for zinc binding.

C:Genetics:

A:Inserts: 40/3; 74/2; 106/3; 136/2; 161/1; 190/1; 266/3; 284/3

C:Superfamily: transcription factor IIIA

C:Keywords: DNA binding; duplication; nucleus; tandem repeat; transcription factor; 2

F:15-37/Region: zinc finger CCH motif

F:45-67/Region: zinc finger CCH motif

F:75-98/Region: zinc finger CCH motif

F:107-129/Region: zinc finger CCH motif

F:137-159/Region: zinc finger CCH motif

F:164-188/Region: zinc finger CCH motif

F:194-214/Region: zinc finger CCH motif

F:223-246/Region: zinc finger CCH motif

F:254-276/Region: zinc finger CCH motif

F:284-304/Region: zinc finger CCH motif

F:312-334/Region: zinc finger CCH motif

F:344-365/Region: zinc finger CCH motif

F:376-398/Region: zinc finger CCH motif

F:409-431/Region: zinc finger CCH motif

F:442-464/Region: zinc finger CCH motif

F:475-497/Region: zinc finger CCH motif

F:508-530/Region: zinc finger CCH motif

F:541-563/Region: zinc finger CCH motif

F:574-596/Region: zinc finger CCH motif

F:607-629/Region: zinc finger CCH motif

F:640-662/Region: zinc finger CCH motif

F:673-695/Region: zinc finger CCH motif

F:706-728/Region: zinc finger CCH motif

F:739-761/Region: zinc finger CCH motif

F:772-794/Region: zinc finger CCH motif

F:805-827/Region: zinc finger CCH motif

F:838-860/Region: zinc finger CCH motif

F:871-893/Region: zinc finger CCH motif

F:904-926/Region: zinc finger CCH motif

F:937-959/Region: zinc finger CCH motif

F:970-992/Region: zinc finger CCH motif

F:1003-1025/Region: zinc finger CCH motif

F:1036-1058/Region: zinc finger CCH motif

F:1069-1091/Region: zinc finger CCH motif

F:1102-1124/Region: zinc finger CCH motif

F:1135-1157/Region: zinc finger CCH motif

F:1168-1190/Region: zinc finger CCH motif

F:1201-1223/Region: zinc finger CCH motif

Query Match 52.4%; Score 1051.5; DB 1; Length 344;

Best Local Similarity 58.1%; Pred. No. 2,2e-72;

Matches 190; Conservative 44; Mismatches 86; Indels 5; Gaps 3;

QY 34 PALPRFICSPDGSANYSKAMKLDALCKHTGERPFVCDYECGCAFRIDYHLSRHLT 93
DB 7 PVAUYKRYICSPADGAGANYNKMKLQHLCKHTGERKPFCEGCKGFTSLHNLTRHSLT 66
QY 94 HTGERPVCAATGCDOKFNTKSLKHPKKNHNOQKQYICSFECCKTKFKKHQDLKHQ 153
DB 67 HTGERPVCAATGCDOKFNTKSLKHPKKNHNOQKQYICSFECCKTKFKKHQDLKHQ 126
QY 154 CQHTNEPLFKCTGECGCKHSPSKLRNAKANEVYCK - GCSFVAKTTELKKNVRE 211
DB 127 FSHQQLPYECRPHGCKDRSLPSLRKHEKVNAGYCKKDDSCSFVCKTTLTKKNVAE 166
QY 212 THKEEITCEYCKRTFKKQYTLQNHKMTAPREVDYCRPRGCGRTYTVFNLQSHLSFH 271
DB 187 CHODLAACDVCKNRKFRKDYLRDQHTHEKERTYVLCSPDGGDRYTTAFMLRSHIOSFH 246
QY 272 EESRPFVCAHAGCKTFAKQSLTRAAVNDPDKKMKLKYKSEKRSLSHLSGY1P 331
DB 247 EQRPFVCAHAGCKTFAKQSLTRAAVNDPDKKMKLKYKSEKRSLSHLSGY1P 304
QY 332 KRKQGGSLSCQNGESPNCVCKMLST 358
DB 305 KSKK-KNAAVSGTEKTDLSLVKKNRPSGT 330

RESULT 4

J01441

transcription factor IIIA - northern leopard frog
C:Species: Rana pipiens (northern leopard frog)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C:Accession: JCI141
R:Gaskins, C.J.; Smith, J.F.; Ogilvie, M.K.; Hanas, J.S.
Gene 120, 197-206, 1992
A>Title: Comparison of the sequence and structure of transcription factor IIIA from Bufo terrestris and Rana pipiens

A:Reference number: JCI141; MUID:99013035; PMID:1398134
A:Accession: JCI141
A:Molecule type: mRNA
A:Residues: 1-335 <GAS>
A:Cross-references: EMBL:X58369; NID:g429060; PIDN:CAA1260.1; PID:g429061
C:Genetics:
A:Gene: TFIIIA
S:Superfamily: Transcription factor IIIA
C:Keywords: DNA binding; duplication; nucleus; transcription factor; zinc finger

Query Match 52.1%; Score 1047; DB 2; Length 335;
Best Local Similarity 62.1%; Pred. No. 4,7e-12;
Matches 187; Conservative 32; Mismatches 76; Indels 6; Gaps 2;

Qy 34 PALPRPICFPSPSSANYSKAWKLDAHLCKHTGHEPFVCDYEGCGKAFFIRDIHLSRIILT 93
||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 PAVYKFVICSFADCSASYNNKKMLDLALCKHTGHEPFCFYVEGGCKGFVTLFPLTHSMIT 66
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 94 HTGKRPVCACATGCDCKDFNTKSNLKKHEBKHNDOOKYICISFECDCKTKRKNOQLIKQ 153
|||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 HTGKRPKCAAPPCDLSFTTWTNLKRNHYORALSLIIECYFDADCOTFKRKNOLKIQ 126
|||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 154 QAHNEDELFKCTOEGCGKHFAPSKILRHAKANEHVYCQ--GCSEPAKWTFELKHYRE 211
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 YIHTNPFCFNCEGDCKSFSRSLKREKHNAAGYPCQKDSCEFPVKTWTEYMKHLA 186
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 212 THKEELLECYCRKTEFRKDYIKOMKHTAPEEDVOCGRBECCGTFTVTYNLOSHLSFH 271
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 SHSEPPTICDVGNKRFKMTKLDKDKRTHEVEREYVWKCPDRSDCDTPTKKGLOSHTSRH 246
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 272 EESRPVFCEHAAGCKTFAMKOSTLRHAVVDPPKKMKMLKYKSKREKRSLSASHLSGYIPP 331
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 EDSPRFACGHGCGKTFAMKOSTLRHANTHDPEKKM----KKRPKPKSLASRLSGYNPK 302
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 332 | 332
Db 303 K 303

RESULT 5
B34895 transcription factor IIIA - Kenyan clawed frog
C:Species: Xenopus borealis (Kenyan clawed frog)
C>Date: 20-Jul-1990 #sequence_revision 21-Oct-1992 #text_change 26-Aug-1999
C:Accession: B34895; S09653
R:John, K.E.; Darby, M.K.; Crawford, E.T.; Brown, D.D.
Cell 61, 293-300, 1990
A>Title: A finger protein structurally similar to TFIIIA that binds exclusively to 5S RNA
A:Reference number: A34895; MUID:90235278; PMID:2331751
A:Accession: B34895
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-339 <TOH>
A:Cross-references: GB:M32472; NID:g214150; PIDN:AAA9713.1; PID:g214151
R:Gaskins, C.J.; Hanas, J.S.
Nucleic Acids Res. 18, 2117-2123, 1990
A>Title: Sequence variation in transcription factor IIYA.
A:Reference number: S09653; MUID:90245658; PMID:2110661
A:Accession: S09653
A:Molecule type: mRNA
A:Residues: 1-221, 'C', 223-234, 'E', 236-291, 'R', 293-309, 'S', 311-313, 'A', 315-318, 'G', 320-339
A:Cross-references: EMBL:M17695; NID:g64474; PIDN:CAA55689.1; PID:g64475
C:Keywords: DNA binding; nucleus; transcription factor; zinc finger

Query Match 51.3%; Score 1031; DB 2; Length 339;

[illegible]

Db 302 KSKKKKKPSOTPFAMESOEQOPDASKADPLVLENLTL 338

RESULT 7

C34895

5S RNA-binding protein p43 - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 20-Jul-1990 #sequence_revision 21-Oct-1992 #text_change 21-Aug-1998

C:Accession: C34895

R:Joho, K.E., Darby, M.K.: Crawford, E.T.; Brown, D.D.

Cell 61, 293-300, 1990

A:Title: A finger protein structurally similar to TFIITa that binds exclusively to 5S RNA

A:Reference number: A34895; MUID:90235278; PMID:2331751

A:Accession: C34895

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: RNA

A:Residues: 1-365 <30H>

C:Superfamily: transcription factor IIIA

C:Keywords: DNA binding; zinc finger

Query Match 27.7%; Score 556.5; DB 2; Length 365;

Best Local Similarity 38.1%; Pred. No. 7.9e-35;

Matches 106; Conservative 48; Mismatches 121; Indels 3; Gaps 3;

Db 42 CSPPDCSANSKMKLDAHLCKHNGEPFVCDYEGCGKAFIRDYHLNHLTHNGEPFV 101

Db 17 CPAAGCAAFRRKSGKLDHMGHSECKPKKCGIKDCDVFAKKQIILKHYKRLALKKLS 76

Qy 102 CAATGCDQDNFKNSLNKKHFERKHNQOQYICSEDECKTKPKKHQOLKIHCOHNEPL 161

Db 77 CPTAGCKMPTSTKSLSRHKLKYNH-GEAVPLKCFVPGCKRSFRKKRLRLHLSVHSNEPL 135

Qy 162 FKCGECCGKHFAPSPKSLKRNKHAHNGYVCO-KGCSFAKMTTELLKHRETNEELICE 220

Db 136 SYCDVPCGCKSSSVAKLVANOKRNHGYCSTEGCOTYSPWTALQTHVKR-HELELOCA 194

Qy 221 VCRKTERKDYLOKMHKTHAPERDVCRCPRGCGRTYTTVFNLOSHIISFEHSRPRVCE 280

Db 195 ACKRPFKASALRRHAKTHAKKFLDLPCCPRDCDTEFSSVNLNHNANVLCLQTHRCP 254

Qy 281 HAGCGTFAMKOSLTRHAVVHDPDKMKKLYKKSREK 318

Db 255 HSCGTRSFAMRESLRLHYVHDPERKKLKLKTVRGPSK 292

RESULT 8

A34895

5S RNA-binding protein p43 - Kenyan clawed frog

C:Species: Xenopus borealis (Kenyan clawed frog)

C>Date: 20-Jul-1990 #sequence_revision 21-Oct-1992 #text_change 21-Aug-1998

C:Accession: A34895

R:Joho, K.E., Darby, M.K.: Crawford, E.T.; Brown, D.D.

Cell 61, 293-300, 1990

A:Title: A finger protein structurally similar to TFIITa that binds exclusively to 5S RNA

A:Reference number: A34895; MUID:90235278; PMID:2331751

A:Accession: A34895

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: RNA

A:Residues: 1-365 <30H>

C:Superfamily: transcription factor IIIA

C:Keywords: DNA binding; zinc finger

Query Match 27.6%; Score 553.5; DB 2; Length 365;

Best Local Similarity 37.4%; Pred. No. 1.3e-34;

Matches 108; Conservative 48; Mismatches 130; Indels 3; Gaps 3;

Db 31 PPRALRRRFICSPDCSANSYKAMKLDHLCKHNGEPFVCDYEGCGKAFIRDYHLNHL 90

Db 6 PTESSKSOVFCSPAAGCAAYRRKGLRDHMGHSECKPKKCGIKDCDVFAKKQIILKHYKRLALKKLS 65

Qy 91 ILTHGKRPFCATGCDQDNFKNSLNKKHFERKHNQOQYICSEDECKTKPKKHQOLK 150

Db 66 MKRHLLTKKHSCTPAAGCKMPTSTKSLSRHKLKYNH-GEAVPLKCFVPGCKRSFRKKRLRL 124

Qy 151 IHOCQHNPLEFCKTOEGCGKHFAPSPKSLKRNKHAHNGYVCO-KGCSFAKMTTELLKH 209

Db 125 IHVSEHSENPVLCVDPVCGCKMSTSAKLAHNRHNGYCRSYDCGTVSPWTALQTHL 184

Qy 210 RETHKEELICEVCKTKFKRKDYLOKMHKTHAPERDVCRCPRGCGRTYTTVFNLOSHIIS 269

Db 185 KK-HPLELOCAACKKPKKASALRRHAKTHAKKFLDLPCCPRDCDTEFSSVNLNHNANVLCLQTHRCP 243

Qy 270 FHESRPFVCEHAGCGTFAMKOSLTRHAVVHDPDKMKKLYKKSREK 318

Db 244 VHLCLOTHRCPSHNCSTRFAMRESLRLHYVHDPERKKLKLKTVRGPSK 292

RESULT 9

337676

zinc-finger protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000

C:Accession: T37676

R:McDougal, R.C.; Rajandream, M.A.; Bartell, B.G.; Davis, P.; Churcher, C.M.

submitted to the EMBL Data Library, October 1999

A:Reference number: Z21736

A:Accession: T37676

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-374 <MCD>

A:Cross-references: EMBL:AL132675; PDB:CA859689.1; GSPDB:GN00066; SPDB:SPAC144.09c

A:Experimental source: strain 972h; cosmid c144

C:Genetics:

A:Gene: SPDB:SPAC144.09c

A:Map position: 1

A:Insertion: 50/3

C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 24.7%; Score 496; DB 2; Length 374;

Best Local Similarity 33.2%; Pred. No. 3.1e-30;

Matches 109; Conservative 48; Mismatches 123; Indels 48; Gaps 12;

Db 40 FICSPDCSANSKMKLDAHLCKHNGEPFVCDYEGCGKAFIRDYHLNHLTHNGEKP 99

Db 23 FHCYERCGKYSRPSLBEQHLRTHSNERPFCVDTGCKAFYRKSHLIKIRCHTVKRP 82

Qy 100 FVCAATGCDQDNFKNSLNKKHFERKHNQOQYICSEDECKTKPKKHQOLKIH-QCOHTN 158

Db 83 FSCYDGDQDNFTQHLNKHTE-VNKKRPACTWAGCDGDFSKHQLSHSACHTH 140

Qy 159 EPLFKCTQEGCGKHFAPSPKSLKRNKHAHNGYVCO-KGCSFAKMTTELLKH 209

Db 141 LLRPCTYQDCELEFATKQLOHNVNRAHEKLIISYCSPEHSCVHGEP--EKWSOLONH 198

Qy 210 RETHKEELICEVCKTKFKRKDYLOKMHKTHAPERDVCRCPRGCGRTYTTVFNLOSH 266

Db 199 REAHVPS--CSICROKFTAAHRLHNVYLHQTLLERKTYHRCPMGSCAKSTRSSALKKH 256

Qy 267 ILSFHEESRPFVCEHAGCGTFAMKOSLTRH-----AVV 300

Db 257 ISVIEGMAFHD--SCGTFKGYKHLQRLHLEGTCKRAHNPYINEGIRHDGEGVAI 314

Qy 301 HDPDKKKMKKLYKKSREKRSLSASHLGY 328

Db 315 HDQKEKELSNVSDVAKK--LINEVIGH 341

RESULT 10

506578

finger protein (clone X1CGF57-1) - African clawed frog (fragment)

C:Species: Xenopus laevis (African clawed frog)

C>Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 31-Dec-1993

C:Accession: S06578

R:Nietfeld, W.; El-Baradi, T.; Mantzel, H.; Pleier, T.; Koester, M.; Poeling, A.; Kno

J. Mol. Biol. 208, 639-659, 1989

A:Title: Second-order repeats in Xenopus laevis finger proteins.

OY 316 REKSLASHSGYIPPRKOGGSLCONGESP-----NCVEDKMLST 358
Db 678 GKAFKLSSTLSTH-----KLIHTGKPYKCEKCGKAFNRPNSNLIHKKIHT 723

Search completed: February 10, 2003, 17:48:19
Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2003, 14:48:36 : Search time 81 Seconds
(Without alignments)
928.484 Million cell updates/sec

Title: US-09-831-426c-2

Perfect score: 2008

Sequence: 1 MDPVVAEVSYSLLTADAF.....ESPNCVEDKMLSTFAVLFLG 365

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.rvirus:*
16: sp.bacteriopl:*
17: sp.archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1583.5	78.8	387	11	09CSH8
2	1582.5	78.8	400	11	08VHT7
3	1498	74.6	336	11	08VHT8
4	946.5	47.1	266	11	09CSW1
5	551.5	27.5	368	13	09W7D7
6	496	24.7	374	3	09UTL5
7	425	21.2	927	4	09UEG4
8	416.5	20.7	927	11	09R161
9	414	20.6	411	13	P70043
10	406.5	20.2	403	5	09YX84
11	406.5	20.2	559	11	09IWM0
12	406.5	20.2	387	11	09Z2X6
13	406.5	20.2	700	4	0961R4
14	405	20.2	411	13	P70046
15	405	20.2	847	4	09H559
16	404	20.1	531	4	09H176

17	403	20.1	934	5	09VKF1
18	402.5	20.0	756	5	061362
19	399	19.9	754	11	099J65
20	399	19.9	962	5	061360
21	398	19.8	891	5	061361
22	397.5	19.8	591	13	090946
23	397.5	19.8	626	4	08W094
24	397.5	19.8	720	6	09N003
25	396	19.7	435	4	096BM3
26	395	19.7	686	4	09MR94
27	395	19.7	803	4	09ETB6
28	393	19.6	751	4	09NS43
29	392.5	19.5	537	4	096OH6
30	392.5	19.5	647	4	096IR2
31	392.5	19.5	715	4	09UEX4
32	392	19.5	626	11	070230
33	392	19.5	846	11	09R163
34	390.5	19.4	346	11	062512
35	390.5	19.4	568	4	08TD23
36	390.5	19.4	576	4	096N22
37	389	19.4	744	11	062788
38	387	19.3	650	11	062886
39	386.5	19.2	873	6	09XS81
40	385	19.2	578	4	096NH0
41	385.5	19.2	569	4	096ND8
42	385.5	19.2	595	4	08FF32
43	385	19.2	600	13	091853
44	384.5	19.1	303	4	09BY31
45	384.5	19.1	357	4	096JW6

ALIGNMENTS

RESULT 1
ID 09CSH8 PRELIMINARY: PRT: 387 AA.
AC 09CSH8
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 2010015D03Rik protein (Fragment).
GN GTF3A OR 2010015D03R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085650; PubMed=11217851;
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashbuner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schral L.M., Scudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gusticich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momaberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wertz C., Whitaker C., Wilming L.,
RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohutsu S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK012977; BAB28476.1; -
HSP: P03001; ITP3.

Query Match	78.9%	Score 1583.5	DR 11	Length 387
Best Local Similarity	77.5%	Pred. No. 8.7e-139		
Matches 283	Conservative 32	Mismatches 47	Indels 3	Gaps 2
QY	1	MDPAAVVAESSVSLTTADAFIAGESSAPTPRPALPRFTICSPFDCSANYSKAMKLDH	60	
DB	24	LEPRVSAEAVSSLTITADAFVGCSE--GAPRPRLPRFTICSPFDCSASANKAMKLDH	81	
QY	61	LCKITGRRPVCYECGCAFIIRDYLSLITHTGKRPFCVCAATGGDOKPNTKSLKH	120	
DB	82	LCKITGRRPVCYECGCAFIIRDYLSLITHTGKRPFCVCAATGGDOKPNTKSLKH	141	
QY	121	FERRHNOQOYICSFEDCKTFFKHQOLIHQCOHTNEPLFKCTQEGCGKHPASPSKL	180	
DB	142	IERKHNPQOYOCYSYEGCKKAFKKHQDLRTHQCOHTNEPLFKCTHCGCGKHPASPSKL	201	
QY	181	RHAKHNGVYCGKGSFVARTWELKHVREPKREILCEVCRTPFKRDLKQHKMTH	240	
DB	202	RHGVHNEGYLCOKGCSMCKTWTLLKHNREAKKEDITCVCGQMFRRDYLNQHKMTH	261	
QY	241	PERDYCACPBBGCGRTYTVFNILQSHIILSFHESSRPVCGHAGCGKTPAMKOSTIRHVV	300	
DB	262	PERDYCACPBBGCGRTYTVFNILQSHIILSFHEKRPVCEHAGCGKTPAMKOSTIRHVV	321	
QY	301	HDPKKKKKLLKVKRSKRSLASHLSGYTPPKKQGGCLSLCQ--NGESPNCVEDKMLTV	359	
DB	322	HDPKKKKKLLKVAAPRRRSLASRLSGYPPPKKQDEPPYSILPKNASSESSSPLEAQLPPA	381	
QY	360	AVLTL 364		
DB	382	ALTLV 386		
RESULT 2				
OBVHT7		PRELIMINARY:	PRT:	400 AA.
ID	QVHT7			
AC	QVHT7			
DT	01-MAR-2002 (TREMblrel, 20, Created)			
DT	01-MAR-2002 (TREMblrel, 20, Last sequence update)			
DT	01-JUN-2002 (TREMblrel, 21, last annotation update)			
DE	Transcription factor IIfa (Fragment).			
GN	GTF3A.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21673987; PubMed=11814676;			
RA	Hanas J.S., Hocker J.R., Cheng Y.-G., Lerner M.R., Brackett D.J.,			
RA	Lightfoot S.A., Hanas R.J., Madhusudan K.T., Moreland R.J.;			
RT	"cDNA cloning, DNA binding, and evolution of mammalian transcription			
RT	factor IIfa.;			
RL	Gene 282;43-52(2002).			
DR	EMBL; AF391799; AAL69686.1; -.			
DR	MCD; MG1;1913846; Gtf3a.			
DR	InterPro; IPRO00822; znf.C2H2.			
DR	PIfam; PFO0096; znf.C2H2; 9.			
DR	SMART; SMO0355; znf.C2H2; 9.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_8.			
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.			
FT	NON_TER 1			
SO	SEQUENCE 400 AA; 45417 MW; EDFEFA8DBD26EAPD CRC64;			

Query Match	78.8%	Score 1582.5	DB 11	Length 400
Best Local Similarity	77.5%	Pred. No. 1,38		
Matches 283	Conservative 32	Mismatches 47	Indels 3	Gaps 2
OY	1	MDPVAIVESVSLTTADAFIAGESSAPTPPPALDPRFICSPDGSANYSKAMKLDAH	60	
DB	37	LEPRVSAVESVSLTTADAFVAGACE--GAPPRALPALSRFICSPDGSASNYKAMKLDAH	94	
OY	61	LCRTGGRPPVPCVCECCGAFIRDVHLSRHLITJTGKEPFVCAATGCDOKFNTKSMKKH	120	
DB	95	LCRTGGRPPVPCVCECCGAFIRDVHLSRHLITJTGKEPFVCAATGCDOKFNTKSMKKH	154	
OY	121	FERHEHMOQOYITCSIEDCKTFPKKQOLKIHCOIHNEPLFKCTQEGCKKHAFASRK	180	
DB	155	IERHNGHPKOYLCSTIEGCKKAFKHQOOLRTHQCOHSEPLFCTHEDGCKKHAFASRK	214	
OY	181	RHAKHNGYVCQCGSFVAKTWTLLKHVRETHKEELICEVCKRTKRYDLKQHKHTA	240	
DB	215	RHGVEHGVLCQCGSFMCQKTWTELLKHREAKHEDITCNVCGMFKRDPYDLQHKHTA	274	
OY	241	PERVYCCPREGCGRTTYTVPNOSHISLTFHESSPPVCEAGCGKTFAMKOSLTPHAY	300	
DB	275	PERVYCCPREGCGRTTYTVPNOSHISLTFHEKRPVCEAGCGKTFAMKOSLTPHAY	334	
OY	301	HDPKKMKMLKVKRSSEKRSLSLHSGYIPPKRQOGSLICQ--NGESPVCVEDKMLSTV	359	
DB	335	HDPKKMKMLKVRAPRERSRLASRLSGYFPKRQEDPDSLPMNASSESSSPPEALLPPA	394	
OY	360	AVLTU 364		
DB	395	ALLTV 399		
RESULT 3				
Q8VHT8	ID	PRELIMINARY:	PRT:	336 AA.
Q8VHT8	Q8VHT8			
AC	01-MAR-2002 (TEMBLrel. 20, Created)			
DT	01-MAR-2002 (TEMBLrel. 20, last sequence update)			
DT	01-JUN-2002 (TEMBLrel. 21, last annotation update)			
DE	Transcription factor I11A (Fragment).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_Taxid:10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN:SPRAGUE-DAWLEY; TISSUE-BRAIN;			
RC	MEDLINE:21673987; Pubmed:1814676;			
RA	Hanas J.S., Hocker J.R., Cheng Y.-G., Lerner M.R., Brackett D.J.,			
RA	"cDNA cloning, DNA binding, and evolution of mammalian transcription			
RT	factor I11A."			
RL	Gene 282:43-52(2002).			
DR	EMBL: AF391798; AAL59685.1;			
DR	InterPro: IPR008022; znf_C2H2.			
DR	Pfam: PF00096; zfc_C2H2_9.			
DR	SMART: SM00355; znf_C2H2_9.			
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_8.			
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.			
FT	NON_TER 1			
FT	NON_TER 1			
SO	SEQUENCE 336 AA; 38514 MW; 7084F95B74D168BA CRC64;			
Query Match	74.6%	Score 1498;	DB 11;	Length 336;
Best Local Similarity	80.3%	Pred. No. 6,4e-131;		
Matches 265	Conservative 23	Mismatches 42;	Indels 0;	Gaps 0;
OY	35	ALPREFICSPDGSANYSKAMKLDAHLCKHCEGEPVDYDGCGKAFIRYHLSRHLLTH	94	
DB	6	ALPREFICSPDGSANYSKAMKLDAHLCKHCEGEPVDYDGCGKAFIRYHLSRHLLTH	65	
OY	95	TGKEPFVCAATGCDOKFNTKSMKKHFERHEHMOQOYICGSEDCCKTFPKKHOOLKIHOC	154	

```

Db 66 TGEPPFVCANNGNOKFSTKSNLKHIERKHENPOKOVNCFNEGCKAFKKHQDLRTHQC 125
Qy 155 QHTNEPLFCTOGCGKHFASPSKLKRNKAHEGYVCGKCSFYAKTTELLKIVRETHK 214
Db 126 QHTNEPLFCTOGCGKHFASPSKLKRNKAHEGYVCGKCSFYAKTTELLKIVRETHK 185
Qy 215 EELICEVCRTEFKRKYLYLKHMTAPREDVRCRDECCGCTTTFVNLQSHLSFHEES 274
Db 186 EELVCTVCGMFRKDKHLKQHMKTAPREDVRCRDECCGCTTTFVNLQSHLSFHEEK 245
Qy 275 RPFVCEHAGCGKTFPANKQSLTRHVVHPDPKKMKLKYKSKREKSLASHLSGYIPPRK 334
Db 246 RPFVCEHAGCGKTFPANKQSLIMRHSVHPDPKKMKLKYKSKREKSLASHLSGYIPPRK 305
Qy 335 OGOSLSCONGSPNCVDEKMLSTVAVLTL 364
Db 306 QEPDSCLPNSTESSSPDEATMLPALTLV 335

```

RESULT 4

Q9CSM1 PRELIMINARY: PRT: 266 AA.

```

AC 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 261011101Rik protein (Fragment).
GN GTF3A OR 261011101Rik.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Hata A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Knehl P., Lewis S., Matsuo Y., Nikolaic I., Persole G., Quackenbush J.,
RA Schirml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Wordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang X.H., Weitz C., Whitaker C., Wilmking L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
DR EMBL: AK011845: BAB27876.1; -.
DR HSSP: P03001: 1TF3.
DR MSD: MG11913846: GL13a.
DR InterPro: IPR000822: ZnF_C2H2.
DR Pfam: PF00096: zf-C2H2; 6.
DR SMART: SM00355: ZnF_C2H2; 6.
DR PROSITE: PS00028: ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS0157: ZINC_FINGER_C2H2_2; 5.
KW DNA-binding; zinc-finger.
FT NON_TER
SQ SEQUENCE 266 AA: 25048 MW: AF2C4738C583518 CMC64;

```

Query Match 47.1%; Score 946.5; DB 11: Length 266;
 Best Local Similarity 77.7%; Pred. No. 7.5e-80;
 Matches 167; Conservative 16; Mismatches 23; Indels 9; Gaps 2;

```

Qy 6 VVABSVSLTADAFIAGESSAPTPPR---PALPRRTGSPFPCSSANYSKAMKLDALHC 62
Db 31 VSPAVASLTIAG-----SSGPVRAAGAPLSPRISCFPCSSASYNKAMKLDALHC 84
Qy 63 KHTGERPFCVDYEGCGKAFIRDYHLRSRLTHTGKRPVCAATGCDCKPNTKSLKKHFE 122
Db 85 KHTGERPFCVDYEGCGKAFIRDYHLRSRLTHTGKRPVCAATGCDCKPNTKSLKKHIE 144
Qy 123 RKHENQOKYITGSEFDCKTFKKHQOLKTHOCQHTNEPLFKTCQDGGCKHFPASPSKLKH 182
Db 145 RKHGNPOKQYVCSYEGCKAFKKHQDLRTHQCQHTNEPLFCTHEGCGKHFASPSKLKH 204
Qy 183 AKAHGYYVCGKCSFYAKTTELLKIVRETHKEE 217
Db 205 GKHVGYLTCGKCSFYAKTTELLKIMREAHKGV 239

```

RESULT 5

Q9W7D7 PRELIMINARY: PRT: 368 AA.

```

AC 09W7D7;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE RNA binding protein 42Sp43.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Belontiiformes; Adiantichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORANGE-RED;
RX MEDLINE=20070874; PubMed=10602271;
RA Kanamori A.;
RT "Systematic identification of genes expressed during early oogenesis
RT in medaka.";
RL Mol. Reprod. Dev. 55:31-36(2000).
CC -1- COFACTOR: BINDS A 2PE-2S CLUSTER (BY SIMILARITY).
DR EMBL: AF128814: MAD38911.2; -.
DR HSSP: P03001: 1TF3.
DR InterPro: IPR000564: 2Fe2S_ferredoxin.
DR InterPro: IPR000822: ZnF_C2H2.
DR Pfam: PF00096: zf-C2H2; 9.
DR SMART: SM00355: ZnF_C2H2; 9.
DR PROSITE: PS00197: 2FE2S_FERREDOXIN; 1.
DR PROSITE: PS00028: ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS0157: ZINC_FINGER_C2H2_2; 6.
KW DNA-binding; Iron-sulfur; Metal-binding; Zinc-finger.
SQ SEQUENCE 368 AA: 42356 MW: A0A84CF04C3B144D CMC64;

```

Query Match 27.5%; Score 551.5; DB 13: Length 368;
 Best Local Similarity 37.9%; Pred. No. 4.9e-43;
 Matches 111; Conservative 36; Mismatches 139; Indels 7; Gaps 4;

```

Qy 40 FTCSFPPCSANYSKAMKLDLKHNGERBPVCDYEGCGKAFIRYHLSRLILHTHEXP 99
Db 16 FTCHAGGACFPREMKLKAHEVHTGERPCACPPAGCGSLPKRSHLKRVHLDHKKVG 75
Qy 100 FVCAATGCDCKPNTKSLKKHFEHENOQKQYITGSEFDCKTFKKHQOLKTHOCQHTNE 159
Db 76 FQCFPANCASFIDAPRLKRNONSAGNH--KFKCQNPQCSLSFKRRLTLKLTKEHNVH 133
Qy 160 PLPFCOTGCGKHFASPSKLKRNKAHEGYV--QKGCSPVAKTWEILLKIVRETHKEEIL 218
Db 134 PNFKCSNICTATPDSIARKAHEKKNAGYSCPHDKCOVVEHTNSKDLRLHAK--HPVSFT 192
Qy 219 CEVCRKTFKRDYLYKHMTAPREDVRCRDECCGCTTTFVNLQSHLSFHEESRPV 278
Db 193 CGVEEKYDRAAGALRRHKRIHASKRPVLLCPRANQAAFTTTFNIEHHIRKVVHLDLLKYK 252
Qy 279 CEHAGCGKTFAMKQSLTRHVVHPDPKKMKLKYKSKREKSLASHLSGYIIP 331

```

```
Db      |   | :|||:::|||   |   |   |   |   |   |   |   |   |   |   |   |
253 CFFPDCRTFVMRESMHRHMVHDDPSAIVK--KQQRPKSMQKRLNGHNQP 302
```

RESULT 6
Q9UTL5
ID Q9UTL5 PRELIMINARY; PRT; 374 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DI 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE ZINC-finger protein (Transcription factor IIfa).
GN SPAC144.09c OR SFC2.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Davis P.,
RA Churcher C.M., Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
RL Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP
RA Schuiteman D.B., Setzer D.R.;
RT "Identification and Characterization of Transcription Factor IIfa from
RT Schizosaccharomyces pombe".
RL Submitted (Mar-2002) to the EMBL/GenBank/DBJ databases.
EMBL: AL132675; CAB59689.1; -.
EMBL: AY091590; AAM00046.1; -.
DR HSSP: P07248; IARD.
DR InterPro: IPR000822; Znf.C2H2.
DR Pfam: PF000056; Zf.C2H2_10.
DR ProDom: PDD00003; Znf.C2H2; 1.
DR SMART: SM00335; Znf.C2H2; 9.
DR PROSITE: PS00026; ZINC_FINGER_C2H2_1; 8.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 9.
DE DNA-binding; Metal-binding; Zinc-finger.
KW Sequencing. 374 AA; 43851 MW; 7469C701FF608FF6 CRC64;


```

DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
DR Zinc finger protein ZFP235.
GN ZFP235.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hauser L., York M., Shannon M., Stubbs L.;
RT "Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7.";
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
CC -i SUBCELLULAR LOCATION: NUCLEAR (By SIMILARITY).
DR EMBL: AF167321; AAD45930.1; -.
DR HSSP: P25490; 12NM.
DR MED: MGI:1929117; zfp235.
DR InterPro: IPR001909; KRAE.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF01352; KRAE; 1.
DR Pfam: PF00096; zf-C2H2; 15.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; znf_C2H2; 12.
DR SMART: SM00349; KRAE; 1.
DR SMART: SM00355; ZNF_C2H2; 15.
DR PROSITE: PSS0805; KRAE; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 15.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 15.
DR DNABinding: Metal-binding; Nuclear protein; Zinc-finger.
KW SEQUENCE 703 AA; 79398 MW; 1A1DA00D10D7386 CRC64;
SQ

```

```

Query Match 20.7%; Score 416.5; DB 11; Length 703;
Best local similarity 33.6%; Pred. No. 3.4e-30;
Matches 115; Conservative 42; Mismatches 142; Indels 43; Gaps 18;

```

```

QY 45 PDCSANSKAMKLDALCKHTGERPVCDEGCGKAFIRDYHLSRHILTHGKERPVCA 104
DB 315 PECKSFQNSHLYAHLPHTGKERPYCCD--NCGKFSSTDLNHCVRHGEKPYKCEV 372
QY 105 TCGDOKFNTKSLKFKERKHNQKQYICSEFDCCKTKRKHQOLKHOCQHTNPLEKC 164
DB 373 --CGKGFQSHLOAH-ERIHNG-EKPKYG--DCKRRSCSNLTHQRVHTGKERPKYC 426
QY 165 TQEGCGHFASPSKLKRAKHAHG---YVCQKCGSVANTTELLKHYR-ETMKEILCE 220
DB 427 DE--CGKRFSLFSLHSHQVRHTEGKERPYCEE--CGKGFSSASFSQSHQVRHTEGKERPVCCS 483
QY 221 VCRKTFKRDYLDKQHMKTAPRDVCRPREGCGRTYTVFNLQSHILSFHEESRPVCE 280
DB 484 VCGKGFSSQSSVFGAHQVRHTEGKERPY-RC--DVCGKRFNMSLNLNH-QRVHTGKERPYCE 539
QY 261 HAGCGKTFAMKQSTIRAAVVDHDDKKMKLKVKKSEKSLASHL-----SGYTPK 332
DB 540 E--CGKGFSSQSSNLQAHQSVHTGEK--PFKNACQKRFSSQSHLOAHQVRHTEGKERPKC 594
QY 333 -----RKQCGSLSCQNSP--NCVE-DKMLSTVAVLT 363
DB 595 DTGKAFSQRNSNLQVHQLIHTGKERPKFCGECGKFSWSAGLT 636

```

```

RESULT 9
P70043 PRELIMINARY; PRT; 411 AA.
ID P70043.
AC P70043.
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Zinc finger protein ZFDL 141.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenocephalia; Xenopus.
OX NCBI_TaxID=8355;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Bourguignon C., Bellefroid E., Bouwmeester T., Rausch O., Pieler T.;
RT "XFDL 141."
RL Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
CC -i SUBCELLULAR LOCATION: NUCLEAR (By SIMILARITY).
DR EMBL: U65897; AAB07010.1; -.
DR HSSP: P25490; 10BD.
DR InterPro: IPR000822; znf_C2H2.
DR Pfam: PF00096; zf-C2H2; 12.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; znf_C2H2; 3.
DR SMART: SM00355; znf_C2H2; 12.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
DR DNABinding: Metal-binding; Nuclear protein; Zinc-finger.
KW SEQUENCE 411 AA; 46615 MW; 0F3984747406A23D CRC64;
SQ

```

```

Query Match 20.6%; Score 414; DB 13; Length 411;
Best local similarity 36.6%; Pred. No. 3.2e-30;
Matches 107; Conservative 28; Mismatches 107; Indels 50; Gaps 15;

```

```

QY 41 ICSPDCSANSKAMKLDALCKHTGERPVCDEGCGKAFIRDYHLSRHILTHGKERPV 100
DB 125 VC--PHCKSTRDXTKTLVHRIHTGTERPVCC--PECGKGFDSFSLASHSHITGKERPV 180
QY 101 VCAATGCDKNTKSLKFKERKHNQKQYIC---SFED----- 138
DB 181 VC--TECGKSFRRKCELSHLO--VHTGVQFVCEGKSFDRSEELNHIHTGKTFF 236
QY 139 ---CKKTFKHHQOLKHOCQHTNPLEKCTQEGCGKHFASPSKLKRAKHAHG---YVC- 191
DB 237 CTECKSKFAAKNKLKRMQIHITGKERPYECTE--CGKQLEKSLKRLNHLSHITGKPVSCF 294
QY 192 QKGSFVAKTTELLKHYRTHKEE--ILCEVCKTFKRDYLDKQHMKTAPRDVCRP 249
DB 295 EGEQF---TWKHLQHLQSHITGERPVCSECGKSFYKASIALHCHITGGERPV-CP 350
QY 250 REGCGRTYTVFNLQSHILSFHEESRPVCEHAQCGCKTFAMKQSTIRAAVVA 301
DB 351 E--CGKGRDPTSLNSHVR-HITGKERPVAC--AECCYTTDSITLKNHLKLAH 397

```

```

RESULT 10
Q9VXB4 PRELIMINARY; PRT; 403 AA.
ID Q9VXB4.
AC Q9VXB4.
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG9609 protein (LD3515Pp).
GN CG9609.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY.
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Burton R.C., Rogers Y.-H.C., Blazey R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-O., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A.I., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

```

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dublin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbagum C.,
 RA Jajalil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Klumel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McInneson D.,
 RA Mekulov G., Mlshina N.V., Moriarty C., Morris J., Mosher D.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamov I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Ventier E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gidbs R.A., Myers E.W., Rubin G.M., Ventier J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2165-2195(2000).
 [2]

SEQUENCE FROM N.A.

STRAIN-BERKELEY

RA Stapleton M., Broksstein P., Hong L., Agbayan A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Flise E., George R.,
 RA Gonzalez M., Garcia H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleib J., Paragas V., Park S., Pounenavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF003503; AAF48663.1; -;
 DR EMBL: AY051849; AAK93273.1; -;
 DR HSSP: P03001; 1TF6;
 DR FLYbase: FBgn0030787; CG9609.
 DR InterPro: IPR000822; znf_C2H2.
 DR Pfam: PF00096; znf_C2H2; 9.
 DR SMART: SM00355; znf_C2H2; 8.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 7.
 DR DNA-binding: Metal-binding; Zinc-finger.
 KW SEQUENCE 403 AA; 46614 MW; 4AD11F430E96662 CRC64;

Query Match 20.2%; Score 406.5; DB 5; Length 403;

Best Local Similarity 30.3%; Pred. No. 1.5e-29;
 Matches 105; Conservative 55; Mismatches 144; Indels 43; Gaps 14;

OY 39 RFGSPDCSANTSKAMKLDLHCKHTGTERPVYCDYEGCGKAFIDYHLSTHRI-LTH--- 94
 DB 34 KYASMPKCEATFRRLDQLDRLREYHHTGILKNACTEGGDKYSLVTLKRLHLSHTHERP 93
 OY 95 --TGEKFPVCAATCCDQKFNFKSLKKEHKKHENDQOQYICSEFCCKTKFKHQOLKI 152
 DB 94 ESAAKTKVCALECSKMFYSVSNMTRHMRTHES-PRKYPCS--QCSAKFSQKLKLRH 150
 OY 153 QC-OHTNEPRLFCKOEGCGKHFASPSKLRHKAHKGYYVCCGCFYAKTTELLKIVRE 211
 DB 151 EIREHTLETPYSCK--CSNGFYQOQOCSHESPSCLTEC-RGCPLOFDKMTLLTKKCRD 207
 OY 212 T-----HKEILCEVCRKTFKRRDYLKQHKMT---HAPERDVC---RCPRREGCGRTYT 258
 DB 208 SLHGKNRHK-----CDRCDSAFDKRPELKRHLLEVKHKEAAQOTDECATSTFCHECGCSYS 263
 OY 259 TVFNLSHLSLTFHESRPFYCEHAGCGKTFPAKKOSTIRIAY-VHPPDKKKKLKLVKKSRE 317
 DB 264 YLRNLKQHMILAH-SGRRFCEQALDGCRCFSSQNLARLLRDLKDHGDGATKELAKKKDK 322
 OY 318 KRSLASHLSGYIPPKKQCGSLCNGESPNCVDEPKMLSTAVVLT 364
 J

DB 323 SKTEGCKTKSTSKRRRQDGRS-----KHSRLSLACIQL 358

RESULT 11

OQ1WMO PRELIMINARY; PRT: 559 AA.

AC OQ1WMO;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical 64.2 kDa protein.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=eye, AND RETINA;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC014712; AAH14712.1; -;
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; znf_C2H2.
 DR Pfam: PF01352; KRAB; 1.
 DR Pfam: PF00096; znf_C2H2; 14.
 DR ProDom: PD000003; znf_C2H2; 6.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN; 14.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 15.
 KW DNA-binding; Hypothetical protein; Zinc-finger.
 SO SEQUENCE 559 AA; 64216 MW; C08F544C0E2A429C CRC64;

Query Match 20.2%; Score 406.5; DB 11; Length 559;

Best Local Similarity 36.2%; Pred. No. 1.2e-29;
 Matches 98; Conservative 42; Mismatches 100; Indels 25; Gaps 14;

OY 40 FICSEPPDCSANTSKAMKLDLHCKHTGTERPVYCDYEGCGKAFIDYHLSTHRIITHTGCKP 99
 DB 279 YVCK--HCGKAYTYNTLRAHRSHTGKRPYCKH--CGKATYSTLRHAHERSHTGCKP 334
 OY 100 FVCAATCGDQKFNFKSLKKEHKKHENDQOQYICSEFCCKTKFKHQOLKIHCOCHTNE 159
 DB 335 YVC--KHCAGKAFQSSYLRLH-KRTHTG-EKPYIC--KCCGKAFARSSHLDIKRSHTGE 386
 OY 160 PLEKCTOEGCGKHFASPSKLRHKAHKG---YVC-QNGCSFVAKTTELLKIVRETHKE 215
 DB 389 KPYVCKQ--CGKAPAOSSYLHIDHRSHTGKRPYVCKQCGKAFTRSSHLDIKH-ITHTGCK 445
 OY 216 EILEVCRKTFKRRDYLKQHKTHAPERD-VCRCPREGCGRTYTVFNLSHLSFHEES 274
 DB 446 PYSCKLCKAFTHSNYLIQIKRHTHGEKRPYCK---ECGKAFASTLSLHIL-EGTISGE 500
 OY 275 RPYVCEHAGCGKTFPAKKOSTIRIAYVHPPDK 305
 DB 501 KPYVCKQ--CGKAFTLSSLRDQVVAHSEK 529

RESULT 12

OQ22X6 PRELIMINARY; PRT: 587 AA.

ID OQ22X6;
 AC OQ22X6;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Zinc finger protein sll-6.
 GN sll-6.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57B6; TISSUE=SPLDN;

RA O-Wang J.: "Zinc finger protein.",
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AB020542; BAA34724.1; -.
DR HSSP: P08047; 1SP2.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF01352; KRAB. 1.
DR Pfam: PF00096; Zf-C2H2. 15.
DR PRINTS: PR00046; ZINC_FINGER.
DR PRODOM: PD000003; ZnF_C2H2. 6.
DR SMART: SM00349; KRAB. 1.
DR SMART: SM00355; ZnF_C2H2. 15.
DR PROSITE: PS50805; KRAB. 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1. 15.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2. 16.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2. 16.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc; Zinc-finger.
SO SEQUENCE 367 AA; 67450 MW; 94B3416511BE9B19 CRC64;

Query Match 20.2%; Score 406.5; DB 11; Length 587;
Best Local Similarity 36.2%; Pred. No. 2.4e-29;
Matches 98; Conservative 42; Mismatches 106; Indels 25; Gaps 14;

OY 40 FICSPDGSANYSAMKIDAHLCKHGEPFVCDYEGCGKAFIRDYHLSRHILHTGTGKRP 99
DB 307 YVCK--HCGKATYTYNTLRHSHSTGKPYVCKH--CGKATYSTLRAHRSHTGKRP 362
OY 100 FVCAATGCDQFNTKSNLKHFERKHENQOYICSEDCCKTFKKHOOIKIHOCQHTNE 159
DB 363 YVC--KHCGKAFGTOSYLRH-KRTHTG--EKPYIC--KCGKAFARSSHLQIHKRSHTGE 416
OY 160 PLFCGTGCGCKHFPASPKLKRHAKAHG---YVC--OKCSFVATWTELLKHVETKE 215
DB 417 KRYVCKO--CGKAFQSSYLHIHQSHSTGKPYVCKGCKGKAFRSHLQIHK--IHTTGEK 473
OY 216 ELICEVCKRTKRDYLDKQHKHTHAPERD--VCRCPREGGRTYTVFNLSHILSPHEES 274
DB 474 PYSCILCKKATLHNSYLDIHKRITGKPYVCK---ECGKAFARSTSLHII--EGTSHGE 528
OY 275 RPEVCHAGCGCKTFAMKOSLTRHAYVHPDK 305
DB 529 KRYVCKO--CGKATLSSLRHSDVHSEK 557

RESULT 13

OY 0961R4 PRELIMINARY; PRT: 700 AA.
AC 0961R4:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to zinc finger protein 224.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Submitted R.;
DR EMBL: BC007303; AA007303.1; -.
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF01352; KRAB. 1.
DR Pfam: PF00096; Zf-C2H2. 19.
DR PRODOM: PD000003; ZnF_C2H2. 5.
DR PROSITE: PS50805; KRAB. 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_18.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 19.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 700 AA; 80559 MW; 79F266EAL8821193 CRC64;

Query Match 20.2%; Score 406.5; DB 4; Length 700;
Best Local Similarity 30.4%; Pred. No. 2.9e-29;
Matches 120; Conservative 45; Mismatches 137; Indels 93; Gaps 19;

OY 46 DCSANYSKAMKIDAHLCKHGEPFVCDYEGCGKAFIRDYHLSRHILHTGTGKPYCAAT 105
DB 236 FCGKGFSSRSLTYVHCKLHSGEKPYNC--EEGKATIHASHIQEHQRIHTGKPKCDT- 292
OY 106 GCDQKENTKSNLKHFERKHENQOYICSEDCCKTFKKHOOIKIHOCQHTNEPLKCT 165
DB 293 -CGKAFRRRSALNNHC--MYHTGKPYKC--EEGCKCFYSSNLRIHQHVHTGKPYKC- 346
OY 166 DCGCKHFPASPKLKRHAKAHG---YVCO---KG-----CSFV 198
DB 347 -EEGCKCIQPSQQAIRIRIHTGKPYVCKGKGYISSSFOAHQGVHTGKPKKNEC 405
OY 199 AKTWTELLKH---VRETHKEELICEVCKRTKRDYLDKQHKHTHAPERDVCRCPRGCG 254
DB 406 GKSFRKTIHYQVHLVHTGKPYKCEYCGKAFQSSYLKIHILKASHVQKPF-KC--EECG 462
OY 255 RYTTVFNLSHILSPHEESRPVCEHAG-----CGKTF 288
DB 463 QGFNQSRLQIHQL-IHTGKPYKCEGCGKGFRRADLIHQRIHTGKPYNCECGKVF 521
OY 289 AMKOSLTRHAYVHPDKKMKLKYKRSKRSLASHL-----SGYIPPKRQ-GGGLS 340
DB 522 SQASHLTHQVHSGEK---PFKCEGCKSFSRSHLQAHQHVHTGKPYKCGCGKGF 578
OY 341 LQGN-----GESP-NCVE-DKMLSTVAVLT 364
DB 579 WSLNLDMHQVHTGKPKPYTGCGCKHFSQSASLQ 613

RESULT 14

ID P70046 PRELIMINARY; PRT: 411 AA.
AC P70046:
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE XFED141.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97321550; Pubmed=9178260;
RA Bellefroid E., Bouguignon C., Bouwmeester T., Rausch O., Blumberg B.,
RA Pieler T.;
RT "Transcription regulation and alternative splicing of an early zygotic
RT gene encoding two structurally distinct zinc finger proteins in
RT Xenopus laevis.";
RL Mech. Dev. 63:99-108(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: U67077; AAB72012.1; -.
DR HSSP: P25490; 10BD.
DR InterPro: IPR000822; ZnF_C2H2.
DR Pfam: PF00096; Zf-C2H2. 12.
DR PRINTS: PR00048; ZINC_FINGER.
DR PRODOM: PD000003; ZnF_C2H2. 3.
DR SMART: SM00355; ZnF_C2H2. 12.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1. 12.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2. 12.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 411 AA; 46530 MW; E7309FEBFB9704 CRC64;

Query Match 20.2%; Score 405; DB 13; Length 411;
Best Local Similarity 35.6%; Pred. No. 2.2e-29;
Matches 104; Conservative 31; Mismatches 107; Indels 50; Gaps 15;

Qy	41	ICSPFDCSANYSKAMLDANLCHTKTGPERPVQYECGGCAFTFDYHLSRHLITHTGKPF	100
Db	125	VC--PHCGKSYBDKTRKLVYHRIYHTGEPFVC--PBCGGQFNDGSLTSHLSTHTGKPF	180
Qy	101	VCAATGCDCKFPTKSNLTKHFERKHEHQOYTC-----SPFD-----	138
Db	181	VC--TECGKSFPRKCELSHQ--YHTGVQAPCTECGKSFPRDSLENIHLNHTGKTFP	236
Qy	139	---CKTPEKKHOOLYHOCQHNPERLCKPOBEGCGAFSPKSLTKRKAAKAGS---VVC-	191
Db	237	CTBCGKSFPAKKNLKRHOMITHTGKPRDCHY--CGQGFLEKSLERHHLSTHTGKPRSCF	294
Qy	192	QKCGSFVAKTYTELLKHYARETHKBE--TLCVCRKTFPKKDYLLKQMKHTAPERDVCRCP	249
Db	295	ECGQGF---TKMHLIDNHOQLSTGTGRPPVSCGCKSLTKTKASALDLCHLHTHTGKREY--CP	350
Qy	250	REGCGRTYTVENLOSHLSTHESRPVCEHAGCGKTFPMQKOSTLTHRAVYH	301
Db	351	E--CGQGFNDPTSLNSVHR--HTGKPRAC--AECCGYTSDSTSLTKNHLKHL	397

RESULT	15
09H559	
ID	PRELIMINARY; PRF: 847 AA.
AC	09H559
DT	01-MAR-2001 (Tremblrel. 16, Created)
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	BA52608.4 (Novel KRAB box containing C2H2 type zinc finger protein (Fragment)).
GN	BA52608.4.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606.
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Babbage A.;
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC	EMBL: AL36981; CAC12728.1; ..
DR	HSSP: P07248.1PAA.
DR	InterPro: IPR001909; KRAB.
DR	InterPro: IPR000822; ZnF_C2H2.
DR	Pfam: PF01352; KRAB.1.
DR	Pfam: PF00096; ZF-C2H2. 17.
DR	PRINTS: PR00048; ZINCINGER.
DR	SMART: SM00349; KRAB.1.
DR	SMART: SM00355; ZnF_C2H2. 17.
DR	PROSITE: PS50805; KRAB.1.
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1. 14.
DR	PROSITE: PS50157; ZINC_FINGER_C2H2_2. 18.
FT	DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
FT	NOV_TER 1 1
CO	SEQUENCE 847 AA: 97660 MW; AB6DC44CDD75541D CRC64;

Query Match	20.28;	Score 405;	DB 4;	Length 847;
Best Local Similarity	32.98;	Pred. No. 4.9e-29;		
Matches 105; conservative	42;	Mismatches 118;	Indels 54;	Gaps 15;

[illegible][illegible]

Search completed: February 10, 2003, 17:47:29
Job time : 89 secs

217 ILCEVCRKTEKRKDY LKOHMKT HAPERDVCRC PRECGRTYTTVFNLSHLSFHEESRP 276


```

XX WPI: 2000-387419/33.
DR P-PSDB: MAY93317.
XX New nucleic acid encoding human transcription factor I11A, useful for
PT treatment and diagnosis of cancer and inherited disease
XX Claim 3: Page 38-39; 49pp; French.
XX
CC The present sequence encodes a human transcription factor designated
CC h11f11A. The polypeptide is probably a DNA-binding protein probably
CC involved in initiating transcription of the gene for ribosomal RNA 5S
CC and maintaining the stability of transcription of other control genes.
CC The h11f11A polynucleotides and polypeptides are used to make
CC therapeutic or diagnostic compositions for diseases associated with
CC disorders of transcriptional control, particularly cancer or other
CC inherited diseases. The h11f11A polynucleotide can also be used to
CC detect anomalies in gene transcription, particularly for diagnosis of
CC inherited disease, also for studying diseases involving h11f11A.
XX
SQ Sequence 1273 BP: 356 A; 329 C; 328 G; 260 T; 0 other;
Query Match 100.0%; Score 1273; DB 21; Length 1273;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGCCGACACGCGGCGCGGACGCGGCGGCTGCTGTACGCGCGCGCTCCCGAAGT 60
DB 1 ATGCCGACACGCGGCGCGGACGCGGCGGCTGCTGTACGCGCGCGCTCCCGAAGT 60
OY 61 GTGCGGCGGCTGCGCGAAGGTTTCAGCAGGAGCGCTGCGCGCGCGCGGCTTCCCGC 120
DB 61 GTGCGGCGGCTGCGCGAAGGTTTCAGCAGGAGCGCTGCGCGCGCGCGGCTTCCCGC 120
OY 121 AGCTGTCTGCGGACGTGGACGCGCGCTGCGCTGCGCTTGGAGGCGCGCGCGCTTGA 180
DB 121 AGCTGTCTGCGGACGTGGACGCGCGCTGCGCTGCGCTTGGAGGCGCGCGCGCTTGA 180
OY 181 TCCGCGGCGCGCTGCGCGAAGGTTTCAGCAGGAGCGCTGCGCGCGCGCTTTCATTC 240
DB 181 TCCGCGGCGCGCTGCGCGAAGGTTTCAGCAGGAGCGCTGCGCGCGCGCTTTCATTC 240
OY 241 AGCGGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCTTCCAGAGGTTATCTG 300
DB 241 AGCGGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCTTCCAGAGGTTATCTG 300
OY 301 CTCTCTTCCCTGACTGACGCGCAATTACAGCAAAAGCCTGGAAGCTTGACGCGCACCTGTG 360
DB 301 CTCTCTTCCCTGACTGACGCGCAATTACAGCAAAAGCCTGGAAGCTTGACGCGCACCTGTG 360
OY 361 CAGGACACGCGGAGAGACCATTTGTTGTGACTATGAAGGTGTGGAGAGGCTTAT 420
DB 361 CAGGACACGCGGAGAGACCATTTGTTGTGACTATGAAGGTGTGGAGAGGCTTAT 420
OY 421 CAGGACTACATCTGAGCCGCCCATCTTCTGACTCAGACAGAGAGAAAACCTTTGTTTG 480
DB 421 CAGGACTACATCTGAGCCGCCCATCTTCTGACTCAGACAGAGAGAAAACCTTTGTTTG 480
OY 481 TCGACCCACTGGCTGTGATCAAAAATTCAACCAAAATCAAACTTGAAGAAACATTTTGA 540
DB 481 TCGACCCACTGGCTGTGATCAAAAATTCAACCAAAATCAAACTTGAAGAAACATTTTGA 540
OY 541 ACGCAAAATCAAAATCAAAACAAATATATATGACGTTTGAAGACTGAAGAGAC 600
DB 541 ACGCAAAATCAAAATCAAAACAAATATATATGACGTTTGAAGACTGAAGAGAC 600
OY 601 CTTTAAAGAAATCAGACGCTGAAATTCATCAGTGGCAGCATACCAATGAAGACCTCTATT 660
DB 601 CTTTAAAGAAATCAGACGCTGAAATTCATCAGTGGCAGCATACCAATGAAGACCTCTATT 660
OY 661 CAAGTGTACCCAGAGAGATGTGGAAACATTGTGATCACCAGCAAGCTGAAGAGACA 720
DB 661 CAAGTGTACCCAGAGAGATGTGGAAACATTGTGATCACCAGCAAGCTGAAGAGACA 720

```

```

OY 721 TCCCAAGCCCCAGAGGCTATGTATGTCAAAAAGATGTTCCCTTTGTGGCAAAACATG 780
DB 721 TCCCAAGCCCCAGAGGCTATGTATGTCAAAAAGATGTTCCCTTTGTGGCAAAACATG 780
OY 781 GACGAACTCTGAAACATGTGAGAGAAACCCATAAGAGAAATACATATGAGATG 840
DB 781 GACGAACTCTGAAACATGTGAGAGAAACCCATAAGAGAAATACATATGAGATG 840
OY 841 CCGGAAACATTTTAAACCCAAATTTACCTTAAGCAACATGAAACTCATGCCCGCA 900
DB 841 CCGGAAACATTTTAAACCCAAATTTACCTTAAGCAACATGAAACTCATGCCCGCA 900
OY 901 AAGGATATATGCTGCTGTCCAGAGAGAGCTGTGGAGAACCTATACTACTGTGTTAA 960
DB 901 AAGGATATATGCTGCTGTCCAGAGAGAGCTGTGGAGAACCTATACTACTGTGTTAA 960
OY 961 TCTCAAAAGCATATCTCTCTCTCCATGAGAAAGCCCGCTTTGTGTGAGAACATGC 1020
DB 961 TCTCAAAAGCATATCTCTCTCTCCATGAGAAAGCCCGCTTTGTGTGAGAACATGC 1020
OY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTACTAGCATGCTGTGTACATGA 1080
DB 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTACTAGCATGCTGTGTACATGA 1080
OY 1081 TCCCTACAAAGAAATGAAGCTCAAAATGTCAAAATAATCTCGTGAAGAAACGAGTTGGC 1140
DB 1081 TCCCTACAAAGAAATGAAGCTCAAAATGTCAAAATAATCTCGTGAAGAAACGAGTTGGC 1140
OY 1141 CTCTCATCTCAGTGTATATATCTCTCCCAAAAGCAAGGCAAGGCTTATCTTTGTG 1200
DB 1141 CTCTCATCTCAGTGTATATATCTCTCCCAAAAGCAAGGCAAGGCTTATCTTTGTG 1200
OY 1201 TCAAAAGGAGAGTACCCCACTGTGGAAGCAAGATCTCTGACAGTTGCACTACT 1260
DB 1201 TCAAAAGGAGAGTACCCCACTGTGGAAGCAAGATCTCTGACAGTTGCACTACT 1260
OY 1261 TACCCCTTGGCTAA 1273
DB 1261 TACCCCTTGGCTAA 1273
RESULT 2
ABK83650
ID ABK83650 standard; cDNA; 1381 BP.
XX
XX ABK83650;
XX
XX 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #221.
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
XX viral infection; parasitic infection; protozoal infection;
XX fungal infection; sterile inflammatory disease; psoriasis;
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;
XX adult respiratory distress syndrome; inflammatory bowel disease;
XX Crohn's disease; ulcerative colitis; periodontal disease;
XX granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US30821.
XX
XX 03-OCT-2000; 2000US-237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX

```

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI: 2002-435328/46.
XX Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
XX drug toxicity -
PS Claim 1: SEQ ID No 221; 114pp: English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GCA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GCA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, AIDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection, and
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp://ipo.int/pub/published_pcl_sequences.
XX
XX Sequence 1381 BP: 386 A: 350 C: 354 G: 291 T: 0 other:
S0

Query Match 96.9%; Score 1233.2; DB 24; Length 1381;
Best Local Similarity 99.5%; Pred. No. 9.6e-312;
Matches 1268; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 1 ATGCGAGCAGGCGGCGGCGGCGGCTGCTGTCAGCCGCGCGCTCCCGAGAGT 60
DB 20 ATGCGAGCAGGCGGCGGCGGCGGCGGCTGCTGTCAGCCGCGCGCTCCCGAGAGT 79
QY 61 GTGCGGCGCTGCGGAGGTTTCAAGAGGAGCGTGGCGGCGGCGGCGGCTTCCCGCGC 120
DB 80 GTGCGGCGCTGCGGAGGTTTCAAGAGGAGCGTGGCGGCGGCGGCGGCTTCCCGCGC 138
QY 121 AGCTGTCTGCGCAGCTGCGAGCGGCGCTGGCGCTTGGAGCGCGCGCGCGCTTGA 180
DB 139 ACCTGTCTGCGCAGCTGCGAGCGGCGCTGGCGCTTGGAGCGCGCGCGCGCTTGA 198
QY 181 TCGCGCGCGCTGCTGCGGAGTCTGCTGCTTACATGAGCGAGCGCTTGAATTC 240
DB 199 TCGCGCGCGCTGCTGCGGAGTCTGCTGCTTACATGAGCGAGCGCTTGAATTC 258
QY 241 AGCGGCGAGCAGCTCCGACCCCGCGCGCGCTTCCAGAGAGTTCAATCTG 300
DB 259 AGCGGCGAGCAGCTCCGACCCCGCGCGCGCTTCCAGAGAGTTCAATCTG 318
QY 301 CTCCTTCCCTGACTGACGCGCAATTTACAGCAAGCCTGGAAGCTTGACGCGCACCCTGTG 360
DB 319 CTCCTTCCCTGACTGACGCGCAATTTACAGCAAGCCTGGAAGCTTGACGCGCACCCTGTG 378
QY 361 CAAGCAGCGGGGAGAGACCATTTGTTTGTGACTATGAAAGGTGGCAAGCCCTTCAT 420
DB 379 CAAGCAGCGGGGAGAGACCATTTGTTTGTGACTATGAAAGGTGGCAAGCCCTTCAT 438
QY 421 CAGGAGTACATCTGACCGCGCAATTTGACTGACACAGAGAAAGCCGTTGTTTG 480
DB 439 CAGGAGTACATCTGACCGCGCAATTTGACTGACACAGAGAAAGCCGTTGTTTG 498
QY 481 TCGAGCACTGCTGCTGATCAAAAATTCAACACAAATCAAACTTGAAACATTTGA 540
DB 499 TCGAGCACTGCTGCTGATCAAAAATTCAACACAAATCAAACTTGAAACATTTGA 558
QY 541 AGCGAATCTGAAATTCACAAACATATATATGACGTTTGAAGACTGTAACAGAC 600
DB 559 AGCGAATCTGAAATTCACAAACATATATATGACGTTTGAAGACTGTAACAGAC 618
QY 601 CTTTAAAGAACATCAGCAGCTGAAATTCATGACGCGACATACCAATGAACCTCTATT 660
DB 619 CTTTAAAGAACATCAGCAGCTGAAATTCATGACGCGACATACCAATGAACCTCTATT 678
QY 661 CAAGTGTACCCAGAGAGATGTGGGAAACCTTTGCATCCCGAGACCTGAACGACA 720
DB 679 CAAGTGTACCCAGAGAGATGTGGGAAACCTTTGCATCCCGAGACCTGAACGACA 738
QY 721 TCGCAAGGCGCCAGAGGCTGTATGTATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 780
DB 739 TCGCAAGGCGCCAGAGGCTGTATGTATGTCAAAAAGATGTTCTTTGTGCAAAAACATG 798
QY 781 GAGCGACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 840
DB 799 GAGCGACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 858
QY 841 CCGGAAACATTTAAACCAAGATTAACCTTAACCAACATGAAACATCATGCCCGAGA 900
DB 859 CCGGAAACATTTAAACCAAGATTAACCTTAACCAACATGAAACATCATGCCCGAGA 918
QY 901 AAGAGTATATGCTGCTGCTGCAAGAGAGGCTGGGAAGCACTATACATGTTTAA 960
DB 919 AAGAGTATATGCTGCTGCTGCAAGAGAGGCTGGGAAGCACTATACATGTTTAA 978
QY 961 TCTCCAAAGCCATATCTCTCTTCATGAGAGAAAGCCGCTTTTGTGTGAACATGC 1020
DB 979 TCTCCAAAGCCATATCTCTCTTCATGAGAGAAAGCCGCTTTTGTGTGAACATGC 1038
QY 1021 TGGCTGTGGCAAAATTTGCAATGAAACAAAGCTCAGTAGCATGCTGTACATGA 1080
DB 1039 TGGCTGTGGCAAAATTTGCAATGAAACAAAGCTCAGTAGCATGCTGTGTACATGA 1098
QY 1081 TCTGCAAGAGAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAANAAC -GGAGTTTGG 1139
DB 1099 TCTGCAAGAGAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAANAACGGGAGTTTGG 1158
QY 1140 CCTTCATCTGAGTATATATATGCTGCAAGAGAGAGGCAAGGCTTATGTTGT 1199
DB 1159 CCTTCATCTGAGTATATATATGCTGCAAGAGAGAGGCAAGGCTTATGTTGT 1217
QY 1200 GTCAAAACGAGAGTACCAACTGTGTGAAGCAAGATGCTCTGACAGTTGAGTAG 1259
DB 1218 GTCAAAACGAGAGTACCAACTGTGTGAAGCAAGATGCTCTGACAGTTGAGTAG 1277
QY 1260 TTACCTTGGCTAA 1273
DB 1278 TTACCTTGGCTAA 1291
RESULT 3
AAT14038
ID AAT14038 standard; cDNA: 1399 BP.

AC AAT14038: .
 XX
 DT 07-JUL-1996 (first entry)
 DE Transcription factor-IIIA gene.
 XX
 KW Human: transcription factor-IIIA; hTFIIIA; DNA binding protein;
 KW ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE;
 KW primer; PCR; polymerase chain reaction; foetal brain;
 KW anchor primer; diagnostic; probe; transcription control;
 KW antitumour; cancer; therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 20..1291
 FT /tag= 3
 FT /product= Transcription factor-IIIA
 FT /note= "Claim 2"
 FT /tag= b
 FT /product= Zinc finger domains
 FT polyA_signal 1363..1368
 FT /tag= c
 XX
 XX EP704526-A1.
 XX
 XX 03-APR-1996.
 XX
 XX 05-SEP-1995: 95EP-0113908.
 XX
 XX 05-SEP-1994: 94JP-0211022.
 XX
 XX (SAKA) OTSUKA PHARM CO LTD.
 XX
 XX Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;
 DR WPI: 1996-173033/18.
 DR P-PSDB: AAR91305.
 XX
 XX Human Transcription Factor III A gene - useful in regulation of
 PT transcription and for diagnosis and treatment of e.g. cancer related
 PT diseases
 XX
 XX Example 1: Page 11-13; 17pp: English.
 XX
 XX The sequence encodes human transcription factor-IIIA (hTFIIIA), a
 CC DNA binding protein with 9 zinc finger domains, which is necessary
 CC for the initiation of 3S RNA gene transcription, binding to an
 CC internal control region of the 5S gene. The coding region
 CC (claimed) is given in AAT14037. A fragment lacking a 5'-portion of
 CC the gene has been isolated from a human foetal brain cDNA library
 CC (OK7-1), and the 5'-portion of the gene has been isolated by
 CC 5'-rapid amplification of cDNA ends using primers H11-R (AAT14039),
 CC H11-E (AAT14040), H11-H (AAT14041) and AP-2 (AAT14042), complementary to
 CC anchor primer AAT14043. Reverse transcription using H-11R is
 CC followed by anchor primer ligation and PCR using AP-2 and H11-E, to
 CC give a full-length cDNA, OK7. The gene and its encoded protein
 CC may be used in diagnosis, identification or therapy of hereditary
 CC diseases such as cancer, or other diseases resulting from abnormal
 CC transcriptional control, and to analyse the mechanisms involved in
 CC their activity.
 XX
 XX Sequence 1399 BP; 405 A; 349 C; 354 G; 291 T; 0 other;

QY	61	GTGCGGGGCTGCGCGGAAGGTTCAGCAGGAGACCGTGGGGCCGGGCGGGTTCGGCG	120
Dp	80	GTGCGGGGCTGCGCGGAAGGTTCAGCAGGAGACCGTGGGGCCGGGCGGGTTCGGCG	138
QY	121	ACGTGTCGTGCGACGTGGACGCGCGCTGGCGTGGGCTTGGAGGGCCCGGCGCTTGA	180
Dp	139	ACGTGTCGTGCGACGTGGACGCGCGCTGGCGTGGGCTTGGAGGGCCCGGCGCTTGA	198
QY	181	TGCGCGGGCGGTGGGTGCGCGAGTGGGTGCTCTTGAACATTCGCGCACCGTTCATTGC	240
Dp	199	TGCGCGGGCGGTGGGTGCGCGAGTGGGTGCTCTTGAACATTCGCGCACCGTTCATTGC	258
QY	241	AGCGCGCGAGACTCTACCTCCGACCGCGCGCGCGCGCGCGCTTCCACGAGGTTTCCTG	300
Dp	259	AGCGCGCGAGACTCTACCTCCGACCGCGCGCGCGCGCGCGCTTCCACGAGGTTTCCTG	318
QY	301	CTCTCTTCCGTGACTGCTACCGCGCAATTACAGCAAAAGCTGGAGAGCTTACCGGACCTGG	360
Dp	319	CTCTCTTCCGTGACTGCTACCGCGCAATTACAGCAAAAGCTGGAGAGCTTACCGGACCTGG	378
QY	361	CAAGCACACGGGGGAGAGACCATTTGTTGTGTGACTATGAGGGGTGGCAAGGCTTCAT	420
Dp	379	CAAGCACACGGGGGAGAGACCATTTGTTGTGTGACTATGAGGGGTGGCAAGGCTTCAT	438
QY	421	CAAGGACTCATCTGACCGCGCCCATCTTGACTACACAGSAGAAAAAGCGTTTGTG	480
Dp	439	CAAGGACTCATCTGACCGCGCCCATCTTGACTACACAGSAGAAAAAGCGTTTGTG	498
QY	481	TGCAGCTACTGGCTGTGATCAAAAATTCAACAAATACTAAGAAACATTGTA	540
Dp	499	TGCAGCTACTGGCTGTGATCAAAAATTCAACAAATACTAAGAAACATTGTA	558
QY	541	AGCCAAACATGAAATAACAAAAACATTTATATGACAGTTTGAACATCTGTAAGAAC	600
Dp	559	AGCCAAACATGAAATAACAAAAACATTTATATGACAGTTTGAACATCTGTAAGAAC	618
QY	601	CTTTAAGAACAATCAGCAGCTGAATAATCCATCAGTCCAGCATACCAATGACCTTATT	660
Dp	619	CTTTAAGAACAATCAGCAGCTGAATAATCCATCAGTCCAGCATACCAATGACCTTATT	678
QY	661	CAAGTGTACCGAGGAAGATGTGGGAACCTTTGATACCCAGCAAGCTGAAAGACA	720
Dp	679	CAAGTGTACCGAGGAAGATGTGGGAACCTTTGATACCCAGCAAGCTGAAAGACA	738
QY	721	TGCGAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTCTTTGTGGCAAAAACATG	780
Dp	739	TGCGAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTCTTTGTGGCAAAAACATG	798
QY	781	GAGGAACTTCTGAACACATGTGAGAGAAACCCATAAAGAGAAATPACTATGTGAAGTATG	840
Dp	799	GAGGAACTTCTGAACACATGTGAGAGAAACCCATAAAGAGAAATPACTATGTGAAGTATG	858
QY	841	CCGGAACAACATTTAAACCCAAAGATTACTTAAAGCACACATGAAAACTATGCCCAAA	900
Dp	859	CCGGAACAACATTTAAACCCAAAGATTACTTAAAGCACACATGAAAACTATGCCCAAA	918
QY	901	AAGGAGATGTATGTCCCTTCCACAGAGAGGCTGTGGAAAGACCTATACTACTGTGTTAA	960
Dp	919	AAGGAGATGTATGTCCCTTCCACAGAGAGGCTGTGGAAAGACCTATACTACTGTGTTAA	978
QY	961	TTCTTCAAAGCATATCTCTCTCTTCCATGTAGGAAAGCGCGCTTTTGTGTGACATGC	1020
Dp	979	TTCTTCAAAGCATATCTCTCTCTTCCATGTAGGAAAGCGCGCTTTTGTGTGACATGC	1038
QY	1021	TGCGCTGTGCAAAACATTTGCAATGAACAAAGTCTCACTAGGAGATCTTGTATCATGA	1080
Dp	1039	TGCGCTGTGCAAAACATTTGCAATGAACAAAGTCTCACTAGGAGATCTTGTATCATGA	1098
QY	1081	TTCTTGAACAAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGAANAAC-GGAGTTTGG	1133
Dp	1099	TTCTTGAACAAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGAANAACGGAGTTTGG	1151

QY 1140 CCTTCATCTCAGTGGATATATCCCTCCAAAAGAAACAAGGCAAGGCTTATCTTGT 1199
 |||||||
 Db 1159 CCTCTATCTCATGTGATGATAT -CTCCCAAAAAGAAACAAGGCTTATCTTGT 1217
 |||||||
 QY 1200 GTCAAAAGGAGATGACCCCACTGTGTGAAAGACAAGATCTCTCTGACAGTTGACATAC 1259
 |||||||
 Db 1218 GTCAAAAGGAGATGACCCCACTGTGTGAAAGACAAGATCTCTGACAGTTGACATAC 1277
 |||||||
 QY 1260 TTACCTTGGCTAA 1273
 |||||||
 Db 1278 TTACCTTGGCTAA 1291
 |||||||

RESULT 4
 AAT14037
 ID AAT14037 standard: cDNA: 1269 BP.
 XX
 AC AAT14037:
 XX
 DT 07-JUL-1996 (first entry)
 XX
 DE Transcription factor-IIIA gene.
 XX
 DE Human: transcription factor-IIIA; hTFIIIA; DNA binding protein;
 KM ribosome; zinc finger; rapid amplification of cDNA ends; 5'-RACE;
 KM primer: PCR: polymerase chain reaction; foetal brain;
 KM anchor primer: diagnostic; probe; transcription control;
 KM antitumor; cancer; therapy; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH misc_feature 336..1115
 FT /tag= a
 FT /product= Zinc finger domains
 FT 1382..1387
 FT /tag= b

EP704526-A1.
 03-APR-1996.
 05-SEP-1995: 95EP-0113908.
 XX
 PR 05-SEP-1994: 94JP-0211022.
 XX
 PA (SAKA) OTSUKA PHARM CO LTD.
 XX
 PI Fujiwara T, Ozaki K, Shimada Y, Shin S, Takeda S;
 XX
 DR WPI: 1996-173033/18.
 DR P-PSDB: AAR91305.
 XX
 PT Human Transcription Factor III A gene - useful in regulation of
 PT transcription and for diagnosis and treatment of e.g. cancer related
 PT diseases
 XX
 PS Claim 2: Page 10-11: 17pp: English.
 XX

The sequence encodes human transcription factor-IIIA (hTFIIIA),
 a DNA binding protein with 9 zinc finger domains, which is necessary
 for the initiation of 5S RNA gene transcription, binding to an
 internal control region of the 5S gene. A fuller cDNA sequence
 with flanking regions is given in AAT14038. A fragment lacking a
 5'-portion of the gene has been isolated from a human foetal brain
 cDNA library (OTK7-1), and the 5'-portion of the gene has been
 isolated by 5'-rapid amplification of cDNA ends using primers H11-R
 (AAT14039), H11-E (AAT14040), H11-H (AAT14041) and Ap-2 (AAT14042).
 CC complementary to anchor primer AAT14043. Reverse transcription using
 CC H11-R is followed by anchor primer ligation and PCR using Ap-2 and
 CC H11-E, to give a full-length cDNA, OTK7. The gene and its encoded
 CC protein may be used in diagnosis, identification or therapy of
 CC hereditary diseases such as cancer, or other diseases resulting from

CC abnormal transcriptional control, and to analyse the mechanisms
 CC involved in their activity.
 XX
 S0 Sequence 1269 BP: 358 A: 324 C: 329 G: 258 T: 0 other:
 Query Match 96.5%; Score 1228.6; DB 17; Length 1269;
 Best Local Similarity 99.4%; Pred. No. 1.5e-110;
 Matches 1264: Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 1 ATGCGAGCAGACGGCGCCGACGCGGGGGGCGTGTGACCGCGCGCTCCGGAAGT 60
 |||||||
 Db 1 ATGCGAGCAGACGGCGCCGACGCGGGGGGCGTGTGACCGCGCGCTCCGGAAGT 60
 |||||||
 QY 61 GTGCGCGCGCTCCGCGGAAGGTTTCAAGAGAGCGCTGTGGCGCGCGCTTCCGCG 120
 |||||||
 Db 61 GTGCGCGCGCTCCGCGGAAGGTTTCAAGAGAGCGCTGTGGCGCGCGCTTCCGCG 119
 |||||||
 QY 121 ACCTGTCTCGCAGCTGGCAGCGCGCTGTGGCGCTTGGAGCGCGCGCTTGGGA 180
 |||||||
 Db 120 ACCTGTCTCGCAGCTGGCAGCGCGCTGTGGCGCTTGGAGCGCGCGCTTGGGA 179
 |||||||
 QY 181 TCCGCGCGCGCTGTGGCGCGAGTGGTGTGCTTGTGACCATGCGCGCGCTTCCG 240
 |||||||
 Db 180 TCCGCGCGCGCTGTGGCGCGAGTGGTGTGCTTGTGACCATGCGCGCGCTTCCG 239
 |||||||
 QY 241 AGCGCGCGAGCTAGCTAGCTCGACCGCGCGCGCGCGCTTCCGAGAGTTTATCTG 300
 |||||||
 Db 240 AGCGCGCGAGCTAGCTAGCTCGACCGCGCGCGCGCGCTTCCGAGAGTTTATCTG 299
 |||||||
 QY 301 CTCTCTCTCTCTGACCTGACGCGCAATTTACAGCAAGCGCTTGAAGCGCGACTGTG 360
 |||||||
 Db 300 CTCTCTCTCTCTGACCTGACGCGCAATTTACAGCAAGCGCTTGAAGCGCGACTGTG 359
 |||||||
 QY 361 CAAGCACACGGGGGAGAGACCATTTTGTGTGATATGAAGGTTGGCAAGGCTTCT 420
 |||||||
 Db 360 CAAGCACACGGGGGAGAGACCATTTTGTGTGATATGAAGGTTGGCAAGGCTTCT 419
 |||||||
 QY 421 CAGGAGCTACATCTGAGCGCGCACATTTCTGACTCACACAGGAAAGACGTTTGTG 480
 |||||||
 Db 420 CAGGAGCTACATCTGAGCGCGCACATTTCTGACTCACACAGGAAAGACGTTTGTG 479
 |||||||
 QY 481 TGCAGCAGCTGTGCTGTGATTCATAAAATTCACACAAATTCATAAGAACTTTTGA 540
 |||||||
 Db 480 TGCAGCAGCTGTGCTGTGATTCATAAAATTCACACAAATTCATAAGAACTTTTGA 539
 |||||||
 QY 541 ACGCAACATGAAATTCACAAACAAATATATATGCACTTTGAAGACTGTAAAGAAC 600
 |||||||
 Db 540 ACGCAACATGAAATTCACAAACAAATATATATGCACTTTGAAGACTGTAAAGAAC 599
 |||||||
 QY 601 CTTTAAAGAAACATCAGCAGATGAAATTCATCAGTGCACAAATTCAGAACTTAT 659
 |||||||
 Db 600 CTTTAAAGAAACATCAGCAGATGAAATTCATCAGTGCACAAATTCAGAACTTAT 659
 |||||||
 QY 661 CAAGTGTACCCAGGAGAGATGTGGAAACACTTGTGATCCACCGCAGAGCTGAAAGACA 720
 |||||||
 Db 660 CAAGTGTACCCAGGAGAGATGTGGAAACACTTGTGATCCACCGCAGAGCTGAAAGACA 719
 |||||||
 QY 721 TGGCAAGGCGCCAGAGGCTATGTATGCAAAAAGATCTCTTGTGGCAAAAACATG 780
 |||||||
 Db 720 TGGCAAGGCGCCAGAGGCTATGTATGCAAAAAGATCTCTTGTGGCAAAAACATG 779
 |||||||
 QY 781 GACGAGACTTCTGAAACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAATATG 840
 |||||||
 Db 780 GACGAGACTTCTGAAACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAATATG 839
 |||||||
 QY 841 CCGGAAACATTTAAACCGAAAGATTTACCTTAACCAACACATGAAACCTATGCCACA 900
 |||||||
 Db 840 CCGGAAACATTTAAACCGAAAGATTTACCTTAACCAACACATGAAACCTATGCCACA 899
 |||||||
 QY 901 AAGGATGTATCTCCCTGTCCAGAGAGGCTGTGGAAGAACCTTACTACTGTGTAA 960
 |||||||
 Db 900 AAGGATGTATCTCCCTGTCCAGAGAGGCTGTGGAAGAACCTTACTACTGTGTAA 959
 |||||||

OY 961 TGTCCAAAGCCATCTCTCTTCATGAGAAAGCCGCTTTTGTGTGTAACATGC 1020
|||||
DB 960 TGTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTTGTGTGTAACATGC 1019
OY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTACATGCTGTGTACATGA 1080
|||||
DB 1020 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTACATGCTGTGTACATGA 1079
OY 1081 TGGCTGTGGCAAAAGAGTCAATGAACAAGCTCTCAGTACATGCTGTGTACATGA 1139
|||||
DB 1080 TGGCTGTGGCAAAAGAGTCAATGAACAAGCTCTCAGTACATGCTGTGTACATGA 1139
OY 1140 CCTCTCATCTCAGTGTATATTCCTCCCAAAAGGAAGGCAAGGCTTATCTTGT 1199
|||||
DB 1140 CCTCTCATCTCAGTGTATATTCCTCCCAAAAGGAAGGCAAGGCTTATCTTGT 1198
OY 1200 GTCAAAAGAGGAGTCAACCTGTGTGAAGCAAGAGCTCTCTGACATGTGACATGC 1259
|||||
DB 1199 GTCAAAAGAGGAGTCAACCTGTGTGAAGCAAGAGCTCTCTGACATGTGACATGC 1258
OY 1260 TTACCTTGGC 1270
|||||
DB 1259 TTACCTTGGC 1269

RESULT 5

AA15407
ID AA15407 standard; DNA; 1213 BP.

AC AA15407;

DT 04-SEP-2000 (first entry)

DE Fragment of DNA encoding a transcription factor designated htfIIIA.

KM Human; transcription factor; htfIIIA; DNA-binding protein;

KW Transcription; ribosomal RNA 5S gene; transcriptional control;

XX cancer; ss.

XX Homo sapiens.

XX MO200028024-A1.

PN 18-MAY-2000.

PD 09-NOV-1999; 99WD-FR02738.

PF 10-NOV-1998; 98FR-0014146.

PR (HMRI) HOECHST MARION ROUSSEL.

PA Bordon-Pallier F, Rocher C;

PI WPI; 2000-387419/33.

DR New nucleic acid encoding human transcription factor IIA, useful for

XX treatment and diagnosis of cancer and inherited disease --

PT Claim 4: Page 42; 49pp; French.

PS The present sequence represents a fragment of a human transcription

XX factor (designated htfIIIA) gene. The polypeptide is probably a

CC DNA-binding protein probably involved in initiating transcription of

CC the gene for ribosomal RNA 5S and maintaining the stability of

CC transcription of other control genes. The htfIIIA polynucleotides and

CC polypeptides are used to make therapeutic or diagnostic compositions

CC for diseases associated with disorders of transcriptional control,

CC particularly cancer or other inherited diseases. The htfIIIA

CC polynucleotide can also be used to detect anomalies in gene

CC transcription, particularly for diagnosis of inherited disease, also

CC for studying diseases involving htfIIIA.

CC Sequence 1213 BP; 349 A; 309 C; 302 G; 253 T; 0 other;

Query Match 95.2%; Score 1211.4; DB 21; Length 1213;
Best Local Similarity 99.9%; Pred. No. 4,5e-306;
Matches 1212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 61 GTGGGGGGGTGGCGGAAGTTGACGAGGAGGAGCCGCGGGGCGCGCGGTTCCGGC 120
|||||
DB 1 GTGGGGGGGGCGCGGAAGTTGACGAGGAGGAGCCGCGGGGCGCGCGGTTCCGGC 60
OY 121 ACCTGTCTGGGACGCTGGGACGCGGCTGCGGCTTGGAGCGCGGCTTGGAG 180
|||||
DB 61 ACCTGTCTGGGACGCTGGGACGCGGCTGCGGCTTGGAGCGCGGCTTGGAG 120
OY 181 TGGCGGGCGCGGTGGGCTGGGAGTGTGTGCTTGGACATGCGGAGCTTCAATGC 240
|||||
DB 121 TGGCGGGCGCGGTGGGCTGGGAGTGTGTGCTTGGACATGCGGAGCTTCAATGC 180
OY 241 ACCGCGGAGAGCTCAGTCCGACCCGCGCGCGCGCTTCCGAGAGTTCAATGC 300
|||||
DB 181 ACCGCGGAGAGCTCAGTCCGACCCGCGCGCGCGCTTCCGAGAGTTCAATGC 240
OY 301 CTCCTTCCCTGACTGACGCGGCAATTTACGAAAGCCTGGAAGCTTGAAGCTGTG 360
|||||
DB 241 CTCCTTCCCTGACTGACGCGGCAATTTACGAAAGCCTGGAAGCTTGAAGCTGTG 300
OY 361 CAAGCACACGGGGGAGAGACATTTTGTGTGACTATGAAGGTGTGGCAGGCTTAT 420
|||||
DB 301 CAAGCACACGGGGGAGAGACATTTTGTGTGACTATGAAGGTGTGGCAGGCTTAT 360
OY 421 CAGGAGTACATCTGAGCCGCGCATTTGACTGACACAGAGAAAGCCGTTGTTTG 480
|||||
DB 361 CAGGAGTACATCTGAGCCGCGCATTTGACTGACACAGAGAAAGCCGTTGTTTG 420
OY 481 TGCAGCCACTGGCTGTATCAAAATTCACACAAATCAACTTGAAGAAATTTTGA 540
|||||
DB 421 TGCAGCCACTGGCTGTATCAAAATTCACACAAATCAACTTGAAGAAATTTTGA 480
OY 541 ACCCAACATGAAATACACAAACATATATATGAGTTTGAAGACTGTGAAGAGC 600
|||||
DB 481 ACCCAACATGAAATACACAAACATATATATGAGTTTGAAGACTGTGAAGAGC 540
OY 601 CTTTAAAGAAACATGACGCTGAATTCATGCTGACATACCAATGAAGCTTAT 660
|||||
DB 541 CTTTAAAGAAACATGACGCTGAATTCATGCTGACATACCAATGAAGCTTAT 600
OY 661 CAAGTGTACCCAGAGAGATGTGGAAACCTTGTGATCCAGCAAGCTTGAAGACA 720
|||||
DB 601 CAAGTGTACCCAGAGAGATGTGGAAACCTTGTGATCCAGCAAGCTTGAAGACA 660
OY 721 TGGCAAGGCCGACGAGGCTATGTATGTCAAAAGAGTGTCTTGTGGCAAAACATG 780
|||||
DB 661 TGGCAAGGCCGACGAGGCTATGTATGTCAAAAGAGTGTCTTGTGGCAAAACATG 720
OY 781 GACGAACTTGTGAACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAAGATG 840
|||||
DB 721 GACGAACTTGTGAACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAAGATG 780
OY 841 CCGGAAACATTTAAAGCAAGATTTACCTTAAGCAACACATTAAGCTTCAATGC 900
|||||
DB 781 CCGGAAACATTTAAAGCAAGATTTACCTTAAGCAACACATTAAGCTTCAATGC 840
OY 901 AAGGATGTATGTGCTGTCCAGAGAGGCTGTGAGAAACCTATACTACTGTATTA 960
|||||
DB 841 AAGGATGTATGTGCTGTCCAGAGAGGCTGTGAGAAACCTATACTACTGTATTA 900
OY 961 TGTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGGAACATGC 1020
|||||
DB 901 TGTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGGAACATGC 960
OY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTACATGCTGTGTACATGA 1080
|||||
DB 961 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTACATGCTGTGTACATGA 1020

YY	1081	TCGACAAACAAACAAAATGAAGCTCAAAATCTCAAAAAATTCGTGTGAAAAACCGAGATTGGC	1140
YY	1081	TCGACAAACAAACAAAATGAAGCTCAAAATCTCAAAAAATTCGTGTGAAAAACCGAGATTGGC	1140
Db	1021	TCCTGACAAAGAAAGAAATGAGCTCAAACTCAAAAAATCTGTGAAAAACGCAATTTGGC	1080
YY	1141	CTCTCATCTCAGTGGATATATATCTCCCTCCCAAAAGGAAAGAGGCGCAAGGCTATCTTTGTG	1200
Db	1081	CTCTCATCTCAGTGGATATATATCTCCCTCCCAAAAGGAAAGAGGCGCAAGGCTATCTTTGTG	1140
YY	1201	TCAAAACGGAGATCAGCCCAACGTGTGGAACAAAGATGCTCTTCGACAGTTGCAGTACT	1260
Db	1141	TCAAAACGGAGATCAGCCCAACGTGTGGAACAAAGATGCTCTTCGACAGTTGCAGTACT	1200
YY	1261	TACCCCTGGCTAA	1273
Db	1201	TACCCCTGGCTAA	1213
RESULT: 6			
AAH34586			
ID	AAH34586	standard: cDNA: 1349 BP.	
XX	AAH34586;		
AC	03-SEP-2001	(first entry)	
XX	03-SEP-2001	(first entry)	
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:1668.		
XX	Human: colon cancer; colon cancer antigen; diagnosis; detection;		
KW	colorectal carcinoma; chromosome 13; ss.		
XX	Homo sapiens.		
OS	MO200122920-A2.		
PN	05-APR-2001.		
XX	28-SEP-2000; 2000MO-US26524.		
PF	29-SEP-1999; 99US-0157137.		
XX	03-NOV-1999; 99US-0163280.		
PR	(HUMA-) HUMAN GENOME SCI INC.		
XX	Ruben SM, Barash SC, Birse CE, Rosen CA;		
PI	WPI: 2001-235357/24.		
XX	P-PSDB: AAG75181.		
DR	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,		
PT	useful for preventing, diagnosing and/or treating colorectal cancers -		
XX	Claim 1; Page 3273-3274; 9803pp: English.		
XX	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon		
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where		
CC	the proteins are collectively known as colon cancer antigens. The colon		
CC	cancer antigens have cytosolic activity and can be used in gene		
CC	therapy and vaccine production. N and P may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate P		
CC	expression. For example, N and P may be used to treat disorders		
CC	associated with decreased expression by rectifying mutations or deletions		
CC	in a patient's genome that affect the activity of P by expressing		
CC	inactive proteins or to supplement the patient's own production of P.		
CC	Additionally, N may be used to produce the colon cancer-associated P,		
CC	by inserting the nucleic acids into a host cell and culturing the cell		
CC	to express the proteins. N and P can be used in the prevention, diagnosis		
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204		
CC	and AAB77789 represent sequences used in the exemplification of the		
CC	present invention.		
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were		
CC	missing at time of publication, meaning no sequences are present for		
CC	SEQ ID NO:1027 to 1052, 7921 and 7922.		
XX			

Seq	Sequence	1349 BP	A: 337	C: 325	G: 285	T: 2	Other:
50	Query Match	94.7%	Score 1206;	DB 22;	Length 1349;		
	Best Local Similarity	99.2%	Pred. No. 1,2e-304;				
	Matches 1212;	Conservative	0;	Mismatches 10;	Indels	0;	Gaps 0;
OY	52 CCGGAGATGTGGCGGCGGTGGCGGCAAGTTTACAGAGGAGCGGTGGGCGCGGCGCGG	111					
DB	11 CCCACGCGTCCGGCGCGCGCGCCGCAAGGTTCAGAGGAGACCGTGGCGCGGCGCGG	70					
OY	112 GTTCCCGGACAGTGTCTCGGCACGTGGACAGGCGCGGTGGCCCTGGCGTTTGAGGCGCGG	171					
DB	71 GTTCCCGGACAGTGTCTCGGCACGTGGACAGGCGCGGTGGCCCTGGCGTTTGAGGCGCGG	130					
OY	172 CGCCCTTGATCGCGCGCGCGGTGGTGGCGGAGTGGTGTGCTGCTTGTGACATTTGCGGAGCG	231					
DB	131 CGCCCTTGATCGCGCGCGCGGTGGTGGCGGAGTGGTGTGCTGCTTGTGACATTTGCGGAGCG	190					
OY	232 GTTTCATTGACAGCGGCGGAGAGGCTCAGCTCCAGCCCGCGCGCGCGCGCTTCCAGGAG	291					
DB	191 GTTTCATTGACAGCGGCGGAGAGGCTCAGCTCCAGCCCGCGCGCGCGCGCTTCCAGGAG	250					
OY	292 GTTCATCTGTCTCTCCCTCGACGTGCGAGGCGCAATTACAGCAAAACCGTGAAGCTTGAAGC	351					
DB	251 GTTCATCTGTCTCTCCCTCGACGTGCGAGGCGCAATTACAGCAAAACCGTGAAGCTTGAAGC	310					
OY	352 GCACCTTGCGAAGCAGCAGCGGCGGAGAGACCAATTGTGTTGTGACTATGAAAGCTGTGGCAA	411					
DB	311 GCACCTTGCGAAGCAGCAGCGGCGGAGAGACCAATTGTGTTGTGACTATGAAAGGTTGGCAA	370					
OY	412 GGCCTTATCAGGAGCATACCATCTGTGGCGCGCACATTGTGACATCAGCAGAGAGAAAGCC	471					
DB	371 GGCCTTATCAGGAGCATACCATCTGTGGCGCGCACATTGTGACATCAGCAGAGAGAAAGCC	430					
OY	472 GTTGTGTTGTGACAGCCAGCTGGCTGTGATCAAAAATTCAACAAATCAAACTTGAAGAA	531					
DB	431 GTTGTGTTGTGACAGCCAGTGGCTGTGATCAAAAATTCAACAAATCAAACTTGAAGAA	490					
OY	532 ACATTTTGAACGCAACATGAAATAATCAACAAACAAATATATATGACGTTTGAAGACAG	591					
DB	491 ACATTTTGAACGCAACATGAAATAATCAACAAACAAATATATATGACGTTTGAAGACAG	550					
OY	592 TTAGAGACCTTTTAAAGAACATCAGAGCTGAAATTCATCAGTGGCCAGCATATCCAAATGA	651					
DB	551 TTAGAGACCTTTTAAAGAACATCAGAGCTGAAATTCATCAGTGGCCAGCATATCCAAATGA	610					
OY	652 ACCGCTATTTCAAGTGTACCCAGGAAGATGTGGGAAACATTTTCATCCACCACGACAAGCT	711					
DB	611 ACCGCTATTTCAAGTGTACCCAGGAAGATGTGGGAAACATTTTCATCCACCACGACAAGCT	670					
OY	712 GAAACGACATGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGAGATGTTCTTTGTGGC	771					
DB	671 GAAACGACATGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGAGATGTTCTTTGTGGC	730					
OY	772 AAAAACATGTAGCAGCACTCTGTGAACATGTGAGACAAACCCATTAAGAGAAATATACATG	831					
DB	731 AAAAACATGTAGCAGCACTCTGTGAACATGTGAGACAAACCCATTAAGAGAAATATACATG	790					
OY	832 TGAAGTATGCCGGAACAAACATTTAAACGCCAAGATTACCTTAAGCACATGAAACTCA	891					
DB	791 TGAAGTATGCCGGAACAAACATTTAAACGCCAAGATTACCTTAAGCACATGAAACTCA	850					
OY	892 TGCCCCAGAAAGAGATATGTATGTGCGGTGTCAGAGAGAGGCTGTGGAAGAACATATACATC	951					
DB	851 TGCCCCAGAAAGAGATATGTATGTGCGGTGTCAGAGAGAGGCTGTGGAAGAACATATACATC	910					
OY	952 TGTGTTTAAATCTCCAAAGGCATATCTCTCTCTCTCTCATGAGAAAGCCGCTTTTGTGTG	1011					
DB	911 TGTGTTTAAATCTCCAAAGGCATATCTCTCTCTCTCATGAGAAAGCCGCTTTTGTGTG	970					
OY	1012 TGAACATGCTGGCGTGTGGCAAAACATTTGACATGAAACAAAGCTCTCACTAGGACATCTGT	1071					
DB	971 TGAACATGCTGGCGTGTGGCAAAACATTTGACATGAAACAAAGCTCTCACTAGGACATCTGT	1030					

QY 1072 TGTACATGATCTCTGACAGAGAAATGAAGCTCAAAAGTCTTCTGTAAGAACG 1131
DB 1031 TGTACATGATCTCTGACAGAGAAATGAAGCTCAAAAGTCTTCTGTAAGAACG 1090
QY 1132 GAGTTGGGCTCTCATCTCACTGATATATCCCTCCCAAGGAAGGCAAGGCTT 1191
DB 1091 GAGTTGGGCTCTCATCTCACTGATATATCCCTCCCAAGGAAGGCAAGGCTT 1150
QY 1192 ACTTTTGTCTCAAAACGAGAGCTACCCCAACTGTGTGGAGAGAAATCTCTCGACACT 1251
DB 1151 ACTTTTGTCTCAAAACGAGAGCTACCCCAACTGTGTGGAGAGAAATCTCTCGACACT 1210
QY 1252 TGCAGTACTTACCTTGGCTAA 1273
DB 1211 TGCAGTACTTACCTTGGCTAA 1232

RESULT 7
AB054923/c
ID AB054923 standard; cDNA: 690 BP.
AC AB054923;
XX
XX 22-AUG-2002 (first entry)
DE Human ovarian antigen HUSYA18 cDNA, SEQ ID NO:803.
XX
XX
XX Human: ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytotoxic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 13q12.3-13.1;
KW gene; ss.
OS Homo sapiens.
XX
XX MO200200677-A1.
PN
XX 03-JAN-2002.
PD
XX 07-JUN-2001; 2001WO-US18569.
PE
XX 07-JUN-2001; 2001WO-US18569.
PR
XX 07-JUN-2001; 2000US-209467P.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Birse CE, Rosen CA:
XX WPI: 2002-147878/19.
DR P-PSDB: ABP41846.
XX
XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
XX Claim 1; SEQ ID NO 803; 2922pp; English.
PS
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and

CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SO Sequence 690 BP; 154 A; 143 C; 145 G; 243 T; 5 other;

Query Match 38.9%; Score 495; DB 24; Length 690;
Best Local Similarity 94.8%; Pred. No. 3.9e-119;
Matches 529; Conservative 2; Mismatches 25; Indels 2; Gaps 2;
QY 716 CGACATGCCAAGGCCACAGAGGCTATGTCACAAAGATGTCCTTGGCAAAA 775
DB 690 CGACATGCCAAGGCCACAGAGGCTATGTCACAAAGATGTCCTTGGCAAAA 633
QY 776 ACATGGAGCAATCTGTAACATGTGAGAGAAACCCATTAAGAGAAATGTAATGTGA 835
DB 632 ACATGGAGCAATCTGTAACATGTGAGAGAAACCCATTAAGAGAGAAATGTAATGTGA 573
QY 836 GTATGCCGGAACATTTAAAGCAAGATTACCTTAAGCAACATGAAATCATGCGC 895
DB 572 GTATGCCGGAACATTTAAAGCAAGATTACCTTAAGCAACATGAAATCATGCGC 513
QY 896 CCAGAAAGGATGTATGCTGCTGCAAGAGAGCGCTGCAAGAACTATGCTGCTG 955
DB 512 CCAGAAAGGATGTATGCTGCTGCAAGAGAGCGCTGCAAGAACTATGCTGCTG 453
QY 956 TTTATCTCCTCAAGCCATATCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGCA 1015
DB 452 TTTATCTCCTCAAGCCATATCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGCA 393
QY 1016 CATGCTGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCACTAGGATGCTGTGA 1075
DB 392 CATGCTGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCACTAGGATGCTGTGA 333
QY 1076 CATGATCTGACACAGAGAAATGAGCTCAAAAGTCAAAATCTCTGTAAGAAAGGAGT 1135
DB 332 CATGATCTGACACAGAGAAATGAGCTCAAAAGTCAAAATCTCTGTAAGAAAGGAGT 273
QY 1136 TTGGCCTCATCTCAGTGTATATGCTCCCAAGAAAGAAACAGGAGGCTATGCT 1195
DB 272 TTGGCCTCATCTCAGTGTATATGCTCCCAAGAAAGAAACAGGAGGCTATGCT 213
QY 1196 TTGTGTCAAAACGAGAGCTACCCCAACTGTGTGAGAGACAAGTCTGTGACAGTTGCA 1255
DB 212 TTGTGTCAAAACGAGAGCTACCCCAACTGTGTGAGAGACAARATGCTGTGACAGTTGCA 153
QY 1256 GTACTTACCTTGGCTAA 1273
DB 152 GTACTTACCTTGGCTAA 135

RESULT 8
ABK44775/c
ID ABK44775 standard; cDNA: 439 BP.
XX

AC ABK44775;
XX
XX 05-JUN-2002 (first entry)
XX
XX cDNA encoding colon tumour protein, SEQ ID No 326.
DE
XX
XX Human: colon tumour; vaccine; colon cancer; immunogenic;
XX immunotherapy; gene: ss.
OS Homo sapiens.
XX
XX WO200212328-A2.
PN
XX
XX 14-FEB-2002.
PD
XX
XX 31-JUL-2001; 2001WO-US24218.
PE
XX
XX 03-AUG-2000; 2000US-223283P.
PR 28-MAR-2001; 2001US-279763P.
PR 29-JUN-2001; 2001US-302051P.
XX
XX (CORI-) CORIXA CORP.
XX
XX
XX King GE, Meagher MJ, Xu J, Secrist H;
PI
XX
XX WPI; 2002-241739/29.
DR
XX
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
XX
XX Claim 1; SEQ ID No 326; 147pp; English.
XX
XX The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
XX
XX Sequence 439 BP; 106 A; 94 C; 99 G; 140 T; 0 other:
SQ
Query Match 32.9%; Score 418.4; DB 24; Length 439;
Best Local Similarity 99.8%; Pred. No. 3.2e-99;
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 854 AAACGCAAGATTACCTTAAGCAACATGAAACTCATGCCAGAAAGGATGTATGT 913
DB 439 AAACGCAAGATTACCTTAAGCAACATGAAACTCATGCCAGAAAGGATGTATGT 380
QY 914 CGCTGCCAAGAGAGGCTGTGGAAGAACCTATCTACTGTGTTTAACTCCAAAGCCAT 973
DB 379 CGCTGCCAAGAGAGGCTGTGGAAGAACCTATCTACTGTGTTTAACTCCAAAGCCAT 320
QY 974 ATCTCTCTCTTCATGAGAAAGCGCCCTTTGTGTAAACATCTGGCTGTGCAAA 1033
DB 319 ATCTCTCTCTTCATGAGAAAGCGCCCTTTGTGTAAACATCTGGCTGTGCAAA 260
QY 1034 ACATTTGCAATAAACAAAGTCTCAGTAGGATCTGTGTACATGATCTGCAGAAG 1093
DB 259 ACATTTGCAATAAACAAAGTCTCAGTAGGATCTGTGTACATGATCTGCAGAAG 200
QY 1094 AAATATGACCTCAAGTCAAAATCTCTGTGAAAGGAGTTGGCTCTCATCTCACT 1153
DB 199 AAATATGACCTCAAGTCAAAATCTCTGTGAAAGGAGTTGGCTCTCATCTCACT 140

QY 1154 GCATATATCCCTCCCAAAAGCAAGGCAAGCTTATCTTGTGTCAAAACGAGAG 1213
DB 139 GGATATATCCCTCCCAAAAGCAAGGCAAGCTTATCTTGTGTCAAAACGAGAG 80
QY 1214 TCACCCAACTGTGTGGAAGACAGATGCTGTGACATGTCATCTACCTTGACCTAA 1273
DB 79 TCACCCAACTGTGTGGAAGACAGATGCTGTGACATGTCATCTACCTTGACCTAA 20
RESULT 9
ABK45138
ID ABK45138 standard; cDNA; 439 BP.
XX
XX
XX ABK45138;
AC
XX
XX 05-JUN-2002 (first entry)
DE
XX
XX cDNA encoding colon tumour protein, SEQ ID No 689.
DE
XX
XX Human: colon tumour; vaccine; colon cancer; immunogenic;
XX immunotherapy; gene: ss.
XX
XX
XX Homo sapiens.
OS
XX
XX WO200212328-A2.
PN
XX
XX 14-FEB-2002.
PD
XX
XX 31-JUL-2001; 2001WO-US24218.
PE
XX
XX 03-AUG-2000; 2000US-223283P.
PR 28-MAR-2001; 2001US-279763P.
PR 29-JUN-2001; 2001US-302051P.
XX
XX
XX (CORI-) CORIXA CORP.
XX
XX
XX King GE, Meagher MJ, Xu J, Secrist H;
PI
XX
XX WPI; 2002-241739/29.
DR
XX
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
XX
XX Claim 1; SEQ ID No 689; 147pp; English.
XX
XX The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
XX
XX Sequence 439 BP; 141 A; 99 C; 93 G; 106 T; 0 other:
SQ
Query Match 32.7%; Score 416.8; DB 24; Length 439;
Best Local Similarity 99.5%; Pred. No. 8.3e-99;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 854 AAACGCAAGATTACCTTAAGCAACATGAAACTCATGCCAGAAAGGATGTATGT 913
DB 1 AAACGCAAGATTACCTTAAGCAACATGAAACTCATGCCAGAAAGGATGTATGT 60
QY 914 CGCTGCCAAGAGAGGCTGTGGAAGAACCTATCTACTGTGTTTAACTCCAAAGCCAT 973

```

DB      |||||
61  CCCTGTCCAGAGAGGCTGTGGAAGAACCTATACAGTGTGTTTATCTCCAAAGCAT 120
OY      |||||
974  ATCTCTCTCCATCAGAGAGAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGGCANA 1033
DB      |||||
121  ATCTCTCTCTCCATCAGAGAGAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGGCANA 180
OY      |||||
1034  ACATTTCGAAAGAAACAGTCTCAGTAGGCATGCTGTGTATCATGATCTTCAGCAAG 1093
DB      |||||
181  ACATTTCGAAAGAAACAGTCTCAGTAGGCATGCTGTGTATCATGATCTTCAGCAAG 240
OY      |||||
1094  AAAATGAAGCTCAAGTCAAAAAATCTGTGTAAGAAACGAGTTGGCTCTCATCTCACT 1153
DB      |||||
241  AAAATGAAGCTCAAGTCAAAAAATCTGTGTAAGAAACGAGTTGGCTCTCATCTCACT 300
OY      |||||
1154  GGAATATATCTCCCAAGAGCAAGGCAAGCCCTATCTTGTGTCAAAACGAGAG 1213
DB      |||||
301  GGATATATCTCCCTCCCAAGAGCAAGGCAAGCCCTATCTTGTGTCAAAACGAGAG 360
OY      |||||
1214  TCACCAACTGTGTGGAAGACAGATGCTCTCGACAGTTGCACTACTTACCTTGGCTAA 1273
DB      |||||
361  TCACCAACTGTGTGGAAGACAGATGCTCTCGACAGTTGCACTACTTACCTTGGCTAA 420

RESULT 10
AB059499 standard; cDNA: 537 BP.
XX
XX  AC  AB059499:
XX
XX  DT  02-AUG-2002 (first entry)
XX
DE  Human colon cancer related nucleotide sequence SEQ ID NO:3194.
XX
XX  KM  Human; colon cancer; cancer; tissue profiling; forensic; mapping;
XX  KM  genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO200229086-A2.
XX
PD  11-APR-2002.
XX
PF  02-OCT-2001; 2001WO-US30732.
XX
XX  PR  02-OCT-2000; 2000US-237271P.
XX
XX  PA  (FARB ) BAYER CORP.
XX
PI  Burgess C, Astle JH, Carroll E, Catino TJ, Dwyerl P, Molino GA;
PI  Thibaultingam A, Lewis ME;
XX
DR  WPI: 2002-426115/45.
XX
XX  PT  New isolated nucleic acid that is differentially expressed in cancer
XX  PT  tissues useful for determining the presence of colon cancer in a cell
XX  PT  or tissue type, and in antisense therapy
XX
PS  Claim 1: Fig 1; 796pp: English.
XX
XX  CC  AB056306 to AB060787 represent isolated nucleic acids (1) differentially
XX  CC  expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
XX  CC  encoded by the AB060776 to AB060787 nucleic acid sequences. (1) can be
XX  CC  used in antisense therapy. An antibody immunoreactive with a polypeptide
XX  CC  encoded by (1) is useful for detecting cancer in a patient sample, and
XX  CC  for detecting the presence or absence of a polynucleotide encoded by a
XX  CC  nucleic acid which hybridises to (1) in a cell. A probe/primer derived
XX  CC  from (1) can be used for determining the presence of a nucleic acid which
XX  CC  hybridises to (1), and for determining the phenotype of cells in a sample
XX  CC  of cells from a patient. (1) is useful for determining the presence of
XX  CC  colon cancer in a cell or tissue type, for determining the presence or
XX  CC  state of other type of cancer, in antisense therapy, to generate
XX  CC  microarrays on a solid surface, to identify a chromosome on which the

```

```

CC  corresponding gene residues, and in tissue profiling, forensics, genetic
CC  analysis, mapping and diagnostic applications. (1) can be used to raise
CC  antibodies, and to screen for peptide analogues and antagonists.
XX
SQ  Sequence 537 BP: 196 A; 116 C; 106 G; 117 T; 2 other:
Query Match      29.4%; Score 374; DB 24; Length 537;
Best local similarity 96.6%; Pred. No. 1,4e-87;
Matches 402; Conservative 0; Mismatches 12; Indels 2; Gaps 2;
OY      |||||
662  AAGTGTACCCAGAGAGATGTGGAAACACTTTGATCAGCCAGCAAGCTGAACGACAT 721
DB      |||||
123  AAAAGTCCAGAGAGATGTGGAAACACTTTGATCAGCCAGCAAGCTGAACGACAT 182
OY      |||||
722  GCCAAGGCCACAGAGGCTATGTATGTCAAAAAGAGTGTCTTTGTGCAAAAACATGG 781
DB      |||||
183  GCCAAGGCCACAGAGGCTATGTATGTCAAAAAGAGTGTCTTTGTGCAAAAACATGG 242
OY      |||||
782  ACGCAACTTGTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGACATATGC 841
DB      |||||
243  ACGCAACTTGTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGACATATGC 302
OY      |||||
842  CGAAACACTTAAACGCAAGATTACCTTAAGCAACATCAATCATGCCCCAGAA 901
DB      |||||
303  CGAAACACTTAAACGCAAGATTACCTTAAGCAACATCAATCATGCCCCAGAA 362
OY      |||||
902  AGGATGTATGTGCTGTGCAAGAGGCTGTGGAAGAACTTACTTCTGTGTTAAT 961
DB      |||||
363  AGGATGTATGTGCTGTGCAAGAGGCTGTGGAAGAACTTACTTCTGTGTTAAT 421
OY      |||||
962  CTCAAAGCATATCTCTCTCTCATGAGAAAGCCGCTTTGTGTGAACATGCT 1021
DB      |||||
422  CTCAAAGCATATCTCTCTCTCATGAGAAAGCCGCTTTTNGTGAACATGCT 481
OY      |||||
1022  GCGTGTGCGCAAAACA-TTTCATATGAACAAAGTCTCAGTACGATGCTGTGTAC 1076
DB      |||||
482  GCGTGTGCGCAAAACA-TTTCATATGAACAAAGTCTCAGTACGATGCTGTGTAC 537

RESULT 11
ABV49231
XX
XX  ID  ABV49231 standard; cDNA: 466 BP.
XX
XX  AC  ABV49231:
XX
XX  DT  17-SEP-2002 (first entry)
XX
XX  DE  Human prostate expression marker cDNA 49222.
XX
XX  KM  Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX  KM  pharmacogenomic marker; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO200160860-A2.
XX
PD  23-AUG-2001.
XX
XX  PF  20-FEB-2001; 2001WO-US05171.
XX
XX  PR  17-FEB-2000; 2000US-183319P.
XX  PR  16-MAR-2000; 2000US-189862P.
XX  PR  25-MAY-2000; 2000US-207454P.
XX  PR  09-JUN-2000; 2000US-211314P.
XX  PR  18-JUL-2000; 2000US-219007P.
XX  PR  13-DEC-2000; 2000US-255281P.
XX
XX  PA  (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX  PI  Schlegel R, Endege WO, Monahan JE;
XX
XX  DR  WPI: 2001-662795/76.
XX

```

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1, Page 9624; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (1) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (1) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 466 BP; 153 A; 92 C; 102 G; 119 T; 0 other;

SO Query Match 23.7%; Score 301.4; DB 23; Length 466;
Best Local Similarity 99.7%; Pred. No. 1.2e-68;
Matches 302; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 737 GGGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGACGGAACCTTGAAA 796
DB 164 GGGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGACGGAACCTTGAAA 223
QY 797 CATGTGAGGAAGCCATTAAGAGGAATACTATGTGAGTATGCCGGAACATTAA 856
DB 224 CATGTGAGGAAGCCATTAAGAGGAATACTATGTGAGTATGCCGGAACATTAA 283
QY 857 CGCAAGATTACCTTAAGCACATGAAACTGATGCCAGAAAGGATGTATGCCG 916
DB 284 CGCAAGATTACCTTAAGCACATGAAACTGATGCCAGAAAGGATGTATGCCG 343
QY 917 TGCCAGAGAGGCGTGTGGAAGACCTATCTGTGTTATCTCCAAAGCCATATC 976
DB 344 TGCCAGAGAGGCGTGTGGAAGACCTATCTGTGTTATCTCCAAAGCCATATC 403
QY 977 CTCTCCCTTCATGAGAAAGCCGCTTTGTGTGTAACATGCTGTGTGCAAAACA 1036
DB 404 CTCTCCCTTCATGAGAAAGCCGCTTTGTGTGTAACATGCTGTGTGCAAAACA 463
QY 1037 TTT 1039
DB 464 TTT 466

RESULT 12
ABV19459
ID ABV19459 standard; cDNA: 425 BP.
XX
XX ABV19459;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 19450.
XX
KW Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
PN WO200160860-A2.
XX
XX 23-AUG-2001.
PD
XX
PF 20-FEB-2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
PI WPI: 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

XX Claim 1, Page 3184; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (1) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (1) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a patient;

CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 425 BP; 142 A; 79 C; 90 G; 114 T; 0 other;

SO Query Match 23.6%; Score 300.4; DB 23; Length 425;
Best Local Similarity 99.7%; Pred. No. 2e-68;
Matches 301; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 737 GGGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGACGGAACCTTGAAA 796
DB 124 GGGTATGATGTCAAAAAGATGTTCTTGTGCAAAAACATGACGGAACCTTGAAA 183
QY 797 CATGTGAGGAAGCCATTAAGAGGAATACTATGTGAGTATGCCGGAACATTAA 856
DB 184 CATGTGAGGAAGCCATTAAGAGGAATACTATGTGAGTATGCCGGAACATTAA 243
QY 857 CGCAAGATTACCTTAAGCACATGAAACTGATGCCAGAAAGGATGTATGCCG 916
DB 244 CGCAAGATTACCTTAAGCACATGAAACTGATGCCAGAAAGGATGTATGCCG 303
QY 917 TGCCAGAGAGGCGTGTGGAAGACCTATCTGTGTTATCTCCAAAGCCATATC 976
DB 304 TGCCAGAGAGGCGTGTGGAAGACCTATCTGTGTTATCTCCAAAGCCATATC 363
QY 977 CTCTCCCTTCATGAGAAAGCCGCTTTGTGTGTAACATGCTGTGTGCAAAACA 1036
DB 364 CTCTCCCTTCATGAGAAAGCCGCTTTGTGTGTAACATGCTGTGTGCAAAACA 423
QY 1037 TT 1038
DB 424 TT 425

RESULT 13
AAH22129/C
ID AAH22129 standard; cDNA: 1401 BP.
XX
XX AAH22129;
AC
XX

DT 20-AUG-2001 (first entry)
 XX Human eukaryotic initiation factor 3 (heif3) encoding cDNA SEQ ID NO:6.
 DE
 XX
 KW Human; eif3; eukaryotic initiation factor; translation initiation factor;
 KM heif3; dendritic cell; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key location/qualifiers
 FT 237..1073
 FT CDS /*tag= a
 FT /product= "eukaryotic initiation factor (heif3)"
 XX
 XX CN128104-A.
 XX
 XX PD 24-JAN-2001.
 XX
 XX PF 27-JUN-2000; 2000CN-0116791.
 XX
 XX PR 27-JUN-2000; 2000CN-0116791.
 XX
 XX (NAME-) NANFANG RES CENT STATE HUMAN GENE GROUP.
 XX
 XX PI Yang Y, Xiao H, Kang B;
 XX
 XX DR WPI: 2001-282654/30.
 XX
 XX DR P-PSDB: AAB98214.
 XX
 XX PT Human translation initiation factor protein and its coding sequence -
 XX
 XX PS Claim 1: Page 17 (disclosure); 20pp; Chinese.
 XX
 CC The present invention describes a human eukaryotic initiation factor
 CC (eif3), designated heif3, which is expressed in human dendritic cells.
 CC heif3 has translation initiation factor activity. Also described in the
 CC present invention are methods for the preparation and detection of the
 CC heif3 protein and nucleotide sequences. The present sequence encodes
 CC heif3, as given in the present invention.
 CC
 XX Sequence 1401 BP; 459 A; 281 C; 306 G; 355 T; 0 other;
 SQ
 Query Match 19.3%; Score 246; DB 22; Length 1401;
 Best local similarity 100.0%; Pred. No. 5.7e-54;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1028 GCGAACAATTTGCAATGAACAAAGTCTCAGCATGCTGTGTACATGATCCTGAC 1087
 DB 1401 GCGAACAATTTGCAATGAACAAAGTCTCAGCATGCTGTGTACATGATCCTGAC 1342
 QY 1088 AAGAACAATGACGCTCAAGTCAAAAATCTGTGAAAAGGAGTTGGCTCTCAT 1147
 DB 1341 AAGAACAATGACGCTCAAGTCAAAAATCTGTGAAAAGGAGTTGGCTCTCAT 1282
 QY 1148 CTCAGTGATATATCCCTCCCAAGGCAAGGCGATTATCTTTGTCTCAAAAC 1207
 DB 1281 CTCAGTGATATATCCCTCCCAAGGCAAGGCGATTATCTTTGTCTCAAAAC 1222
 QY 1208 CGAGACTGCACCACTGTGTGGAAGACAAAGTGTCTGACAGTGAAGTACTTACCTT 1267
 DB 1221 GGAAGTCCCACTGTGTGGAAGACAAAGTGTCTGACAGTGAAGTACTTACCTT 1162
 QY 1268 GGCATA 1273
 DB 1161
 DB 1161 GGCATA 1156
 RESULT 14
 AAS73534
 ID AAS73534 standard; CDNA: 1889 BP.
 XX
 AC AAS73534;
 XX

DT 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #9338.
 DE
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 XX MO200175067-A2.
 XX
 XX PD 11-OCT-2001.
 XX
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX
 XX PR 31-MAR-2000; 2000US-0540217.
 XX
 XX PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX PI Drmanac RT, Liu C, Tang YT;
 XX
 XX DR WPI: 2001-639362/73.
 XX
 XX DR P-PSDB: ABC09347.
 XX
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX PT diagnostics, forensics, gene mapping, identification of mutations
 XX PT responsible for genetic disorders or other traits and to assess
 XX PT biodiversity -
 XX
 XX PS Claim 1: SEQ ID No 9338; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 1889 BP; 488 A; 453 C; 417 G; 531 T; 0 other;
 Query Match 18.5%; Score 236; DB 23; Length 1889;
 Best local similarity 95.5%; Pred. No. 2.7e-51;
 Matches 275; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
 QY 986 CATGAGAAAGCCGCCCTTTGTGTGTAACATGCTGGCTGTGCAAAACATTTGCAAG 1045
 DB 2 CATGAGAAAGCCGCCCTTTGTGTGTAACATGCTGGCTGTGCAAAACATTTGCAAAAT 59
 QY 1046 AAACAAGTCTCAGTACGATGCTGTGTACATGATCTGACAAGAAATGAAGCTC 1105
 DB 60 AAACAAGTCTCAGTACGATGCTGTGTACATGCTGCTGCAAGAAATGAAGCTC 118
 QY 1106 AAAGTCAAAAATCTGCGAAGAAACGAGTTGGCTCTCATCTCAGTGGATATATCCCT 1165
 DB 119 AAAGTCAAAAATCTGCGAAGAAACGAGTTGGCTCTCATCTCAGTGGATATATCCCT 178
 QY 1166 CCCAAAGGAACAAAGGCGAGGCTTATCTTTGTGTCAAAACGAGAGTACCCAACTGT 1225

THIS PAGE BLANK (USPTO)


```
OY 361 CAGGACACGGGGGAGACACATTTGTTGGACTATGAGGGGTGGCAAGCCCTCAT 420
    |||
Db 380 CAGGACACGGGGGAGACACATTTGTTGGACTATGAGGGGTGGCAAGCCCTCAT 439
OY 421 CAGGAGTACATCTGAGCCGCCACATTTGACTACACAGGAAAGCCGTTGTTG 480
    |||
Db 440 CAGGAGTACATCTGAGCCGCCACATTTGACTACACAGGAAAGCCGTTGTTG 499
OY 481 TGCAGCCACTGGCTGTGATCAAAAATTCACACAAAATCAAACTGAAGAAACATTTGA 540
    |||
Db 500 TGCAGCCACTGGCTGTGATCAAAAATTCACACAAAATCAAACTGAAGAAACATTTGA 559
OY 541 ACGCAAAACATGAAAATCAACAAAAAATATATATGAGTTTGAAGACTGTAGAAGAC 600
    |||
Db 560 ACGCAAAACATGAAAATCAACAAAAAATATATATGAGTTTGAAGACTGTAGAAGAC 619
OY 601 CTTTAAGAAACATGAGCGCTGAAATTCATCAGTGGCAGCATACCAATGAACCTCAT 660
    |||
Db 620 CTTTAAGAAACATGAGCGCTGAAATTCATCAGTGGCAGCATACCAATGAACCTCAT 679
OY 661 CAAGTGTACCCAGGAAGATGTGGGAAACATTTGCATCACCCAGCAAGCTGAAAGACA 720
    |||
Db 680 CAAGTGTACCCAGGAAGATGTGGGAAACATTTGCATCACCCAGCAAGCTGAAAGACA 739
OY 721 TGCCAAGGCCCCAGAGGCTATGTATGTCAAAAAGGATGTCTTGTGGCAAAAAACATG 780
    |||
Db 740 TGCCAAGGCCCCAGAGGCTATGTATGTCAAAAAGGATGTCTTGTGGCAAAAAACATG 799
OY 781 GACGGAATCTTGAACATGTGTAGAGAAACCCATAAAGAGAAATATATGTGAAGTATG 840
    |||
Db 800 GACGGAATCTTGAACATGTGTAGAGAAACCCATAAAGAGAAATATATGTGAAGTATG 859
OY 841 CCGGAAACATTTAAACCAAGATTAACCTTAAACCAACATGAAAAACATGCCCCAGA 900
    |||
Db 860 CCGGAAACATTTAAACCAAGATTAACCTTAAACCAACATGAAAAACATGCCCCAGA 919
OY 901 AAGGATGTATGTGCTGCTGCCAAGAGAGGCTGTGGAAGAACTATACTGTGTGTTAA 960
    |||
Db 920 AAGGATGTATGTGCTGCTGCCAAGAGAGGCTGTGGAAGAACTATACTGTGTGTTAA 979
OY 961 TCTCCAAAGCCATATCTCTCTCCATGAGAGAGGCGCCTTTGTGTGGAACATGC 1020
    |||
Db 980 TCTCCAAAGCCATATCTCTCTCCATGAGAGAGGCGCCTTTGTGTGGAACATGC 1039
OY 1021 TGGCTGTGGCAAAATTTGCAATGAAACAAGCTCTACTAGCATGCTGTTGTACATGA 1080
    |||
Db 1040 TGGCTGTGGCAAAATTTGCAATGAAACAAGCTCTACTAGCATGCTGTTGTACATGA 1099
OY 1081 TCGTGACAGAAGAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAAGAAACGAGTTTGGC 1140
    |||
Db 1100 TCGTGACAGAAGAAATGAAGCTCAAAAGTCAAAAAATCTCGTGAAGAAACGAGTTTGGC 1159
OY 1141 CTCTCATCTCATGTGATATATCCCTCCCAAGAGAAACAAGGCAAGCTTATCTTTGTG 1200
    |||
Db 1160 CTCTCATCTCATGTGATATATCCCTCCCAAGAGAAACAAGGCAAGCTTATCTTTGTG 1219
OY 1201 TCAGAAACGAGAGATCCCAACATGTGTGGAAGACAGATGCTCTGACAGTTGACAGTACT 1260
    |||
Db 1220 TCAGAAACGAGAGATCCCAACATGTGTGGAAGACAGATGCTCTGACAGTTGACAGTACT 1279
OY 1261 TACCCCTTGCTAA 1273
    |||
Db 1280 TACCCCTTGCTAA 1292
    |||
```

```
RESULT 2
US-09-724-676A-44689
; Sequence 44689, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
```

```
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: Patent version 3.2
; SEQ ID NO 44689
; LENGTH: 1906
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1814)..(1814)
; OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-44689
```

```
Query Match          99.6%; Score 1268.2; DB 5; Length 1906;
Best local similarity 99.8%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 1 ATGCGCAGCAGCGCGCGCGGCGGCGGCTGCTGTGACCGCGGCTCCCGAAGT 60
    |||
Db 20 ATGCGCAGCAGCGCGCGCGGCGGCGGCGGCTGCTGTGACCGCGGCTCCCGAAGT 79
OY 61 GTGCGGCGCTCGCGCGAAGTTCAAGAGGAGCGCGTGGCGCGCGCGGCTCCCGG 120
    |||
Db 80 GTGCGGCGCGCGCGCGAAGTTCAAGAGGAGCGCGTGGCGCGCGCGGCTCCCGG 139
OY 121 AGGTGTCTGGCAGCTGGCAGCGCGGCTGGCCCTTGAGAGCGCGCGCTTGA 180
    |||
Db 140 AGGTGTCTGGCAGCTGGCAGCGCGGCTGGCCCTTGAGAGCGCGCGCTTGA 199
OY 181 TCGCGGCGCGGTGTGCGCGAGTGGTGTCTGCTGACCATGCGCGAGCTGATTC 240
    |||
Db 200 TCGCGGCGCGGTGTGCGCGAGTGGTGTCTGCTGACCATGCGCGAGCTGATTC 259
OY 241 AGCGCGGAGACTGAGCTCGGACCGCGCGCGCGCGCTTCCAGAGGTTGATCG 300
    |||
Db 260 AGCGCGGAGACTGAGCTCGGACCGCGCGCGCGCGCTTCCAGAGGTTGATCG 319
OY 301 CTCTCTCCCTGACTGACGCGGCAATTACAGCAAGCCTGGAACCTTGACGCGC 360
    |||
Db 320 CTCTCTCCCTGACTGACGCGGCAATTACAGCAAGCCTGGAACCTTGACGCGC 379
OY 361 CAAGCAACGGGGGAGACATTTGTTGAGTATGAAGAGGTTGGGAAGCCCTCAT 420
    |||
Db 380 CAAGCAACGGGGGAGACATTTGTTGAGTATGAAGAGGTTGGGAAGCCCTCAT 439
OY 421 CAGGGACTACATCTGAGCGCGCACTCTGACTCACACAGAGAAACCGTTGTTG 480
    |||
Db 440 CAGGGACTACATCTGAGCGCGCACTCTGACTCACACAGAGAAACCGTTGTTG 499
OY 481 TGCAGCACTGGCTGTGATCAAAAATTCACACAAAATCAAACTGAAGAACTTTGA 540
    |||
Db 500 TGCAGCACTGGCTGTGATCAAAAATTCACACAAAATCAAACTGAAGAACTTTGA 559
OY 541 ACGCAAAACATGAAAATCAACAAAAAATATATATGAGTTTGAAGACTGTAGAAGAC 600
    |||
Db 560 ACGCAAAACATGAAAATCAACAAAAAATATATATGAGTTTGAAGACTGTAGAAGAC 619
OY 601 CTTTAAGAAACATGAGCGCTGAAATTCATCAGTGGCAGCATACCAATGAACCTCAT 660
    |||
Db 620 CTTTAAGAAACATGAGCGCTGAAATTCATCAGTGGCAGCATACCAATGAACCTCAT 679
OY 661 CAAGTGTACCCAGGAAGATGTGGGAAACATTTGCATCACCCAGCAAGCTGAAAGACA 720
    |||
Db 680 CAAGTGTACCCAGGAAGATGTGGGAAACATTTGCATCACCCAGCAAGCTGAAAGACA 739
OY 721 TGCCAAGGCCCCAGAGGCTATGTATGTCAAAAAGGATGTCTTGTGGCAAAAAACATG 780
    |||
Db 740 TGCCAAGGCCCCAGAGGCTATGTATGTCAAAAAGGATGTCTTGTGGCAAAAAACATG 799
OY 781 GACGGAATCTTGAACATGTGTAGAGAAACCCATAAAGAGAAATATATGTGAAGTATG 840
    |||
Db 800 GACGGAATCTTGAACATGTGTAGAGAAACCCATAAAGAGAAATATATGTGAAGTATG 859
```

```

OY 841 CGGAAACATTTAAAGCAAGATTAAGCAACATGAAGAAACTCATGCCCGA 900
    |||
Db 860 CGGAAACATTTAAAGCAAGATTAAGCAACATGAAGAAACTCATGCCCGA 919
OY 901 AAGGATGTATGTGCTGTCTCCAGAGAGGCTGTGGAAGAAGCTATACTACTGTGTAA 960
    |||
Db 920 AAGGATGTATGTGCTGTCTCCAGAGAGGCTGTGGAAGAAGCTATACTACTGTGTAA 979
OY 961 TCTCAAGGACATATCCCTCTCCATGAGAGAGGCGGCTTTGTGTGTGAACATGC 1020
    |||
Db 980 TCTCAAGGACATATCCCTCTCTCTCCATGAGAGAGGCGGCTTTGTGTGTGAACATGC 1039
OY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTAGCATGTGTGTACATGA 1080
    |||
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTAGCATGTGTGTACATGA 1099
OY 1081 TCTGTACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGGAAGAAAGGAGTTTGGC 1140
    |||
Db 1100 TCTGTACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGGAAGAAAGGAGTTTGGC 1159
OY 1141 CTCTCATCTCAGTGTGATATATCCCTCCCAAGAGAAACAAGGCAAGCTTATCTTTGTG 1200
    |||
Db 1160 CTCTCATCTCAGTGTGATATATCCCTCCCAAGAGAAACAAGGCAAGCTTATCTTTGTG 1219
OY 1201 TCATAAGGAGAGTCCACCACTGTGTGGAAGCAAGATGCTGTGACAGTTGCACTACT 1260
    |||
Db 1220 TCATAAGGAGAGTCCACCACTGTGTGGAAGCAAGATGCTGTGACAGTTGCACTACT 1279
OY 1261 TACCTTTGGCTAA 1273
    |||
Db 1280 TACCTTTGGCTAA 1292

RESULT 3
US-09-724-676-44634
: Sequence 44634, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 44634
: LENGTH: 2019
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1927)-(1927)
: OTHER INFORMATION: n is a, c, g, or t
US-09-724-676-44634

Query Match          99.8%; Score 1268.2; DB 5; Length 2019;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAGGCGACAGAGGCGGCGGAGGCGGCGGTGCTGTGACCGCGGCGCTCCCGGAAGT 60
    |||
Db 20 AAGGCGACAGAGGCGGCGGAGCGGCGGCGGTGCTGTGACCGCGGCGCTCCCGGAAGT 79
OY 61 GTGCGGCGGTCCGCGCAAGTTCAGCAGAGGAGCGGTGGGCGGCGCGGCTTCCCGGC 120
    |||
Db 80 GTGCGGCGGCGCGCGCAAGTTCAGCAGAGGAGCGGTGGGCGGCGCGGCTTCCCGGC 139
OY 121 AGGTGTCTGAGCAGTGTGCGAGGCGGCGGCGGCGGCGGCTTGGAGGCGGCGGCGGCGG 180
    |||
Db 140 AGGTGTCTGAGCAGTGTGCGAGGCGGCGGCGGCGGCGGCGGCTTGGAGGCGGCGGCGG 199
OY 181 TCCGCCGCGCGGTGTGCGCGAGTGTGCTGCTTGCACCATGCCGAGCGCTTATTTGC 240
    |||
Db 200 TCCGCCGCGCGGTGTGCGCGAGTGTGCTGCTTGCACCATGCCGAGCGCTTATTTGC 259
```

```

OY 241 AGCGCGGAGAGCTACGTCCGAGCCCGCGCGCCCGGCTTCCAGAGGTTTATCTG 300
    |||
Db 260 AGCGCGGAGAGCTACGTCCGAGCCCGCGCGCCCGGCTTCCAGAGGTTTATCTG 319
OY 301 CTCTTCTCCCTGACGACGCGCAATATACAGCAAGGCTGGAGGCTTGACGCGCACCTGTG 360
    |||
Db 320 CTCTTCTCCCTGACGACGCGCAATATACAGCAAGGCTGGAGGCTTGACGCGCACCTGTG 379
OY 361 CAAGCAGACGCGGAGAGACCATTTTGTGTGACTATGAAGGTGTGCGAAGGCTTAT 420
    |||
Db 380 CAAGCAGACGCGGAGAGACCATTTTGTGTGACTATGAAGGTGTGCGAAGGCTTAT 439
OY 421 CAGGAGTACCATCTGAGCGCGCACATTTGACTGACACAGAGGAAAGCCGTTGTGTG 480
    |||
Db 440 CAGGAGTACCATCTGAGCGCGCACATTTGACTGACACAGAGGAAAGCCGTTGTGTG 499
OY 481 TGCAGCCACTGGCTGTGATTCAAAAATTCAAACAAAAATCAAACTTGAGAAACATTTTGA 540
    |||
Db 500 TGCAGCCACTGGCTGTGATTCAAAAATTCAAACAAAAATCAAACTTGAGAAACATTTTGA 559
OY 541 ACGCAAAACATGAATAATCAACAAATATATATGAGTTTGAAGACTGTAAAGAGC 600
    |||
Db 560 ACGCAAAACATGAATAATCAACAAATATATATGAGTTTGAAGACTGTAAAGAGC 619
OY 601 CTTTAAGAAACATCAGCAGCTGAAATATCATCAGTGCACCATACCATGAAGCTCTATT 660
    |||
Db 620 CTTTAAGAAACATCAGCAGCTGAAATATCATCAGTGCACCATACCATGAAGCTCTATT 679
OY 661 CAAGTGTACCCAGAGAGTGTGGAAACCTTTGATCCACCGACAGCTGAAGAGCA 720
    |||
Db 680 CAAGTGTACCCAGAGAGTGTGGAAACCTTTGATCCACCGACAGCTGAAGAGCA 739
OY 721 TGCCAAGGCGCCACGAGGCTATGTATGTCAAAAAGCATTTCTTTGTGCAAAAACATG 780
    |||
Db 740 TGCCAAGGCGCCACGAGGCTATGTATGTCAAAAAGCATTTCTTTGTGCAAAAACATG 799
OY 781 GAGGAACTTCTGAACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAAGTATG 840
    |||
Db 800 GAGGAACTTCTGAACATGTGAGAGAAACCATTAAGAGAAATCTATGTGAAGTATG 859
OY 841 CGGAAACATTTAAAGCAAGATTAAGCAACATGAAGAAATCTATGTGAAGTATG 900
    |||
Db 860 CGGAAACATTTAAAGCAAGATTAAGCAACATGAAGAAATCTATGTGAAGTATG 919
OY 901 AAGGATGTATGCTGTCTCCAGAGAGGCTGTGGAAGCAACTATACTACTGTGTAA 960
    |||
Db 920 AAGGATGTATGCTGTCTCCAGAGAGGCTGTGGAAGCAACTATACTACTGTGTAA 979
OY 961 TCTCAAGGACATATCTCTCTCTTCATGAGAGAGCGGCGCTTTTGTGTGCAACATGC 1020
    |||
Db 980 TCTCAAGGACATATCTCTCTCTTCATGAGAGAGCGGCGCTTTTGTGTGCAACATGC 1039
OY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTAGCATGTGTGTACATGA 1080
    |||
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTCAGTAGCATGTGTGTACATGA 1099
OY 1081 TCTGTACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGGAAGAAAGGAGTTTGGC 1140
    |||
Db 1100 TCTGTACAAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGGAAGAAAGGAGTTTGGC 1159
OY 1141 CTCTCATCTCAGTGTGATATATCCCTCCCAAGAGAAACAAGGCAAGCTTATCTTTGTG 1200
    |||
Db 1160 CTCTCATCTCAGTGTGATATATCCCTCCCAAGAGAAACAAGGCAAGCTTATCTTTGTG 1219
OY 1201 TCATAAGGAGAGTCAACCACTGTGTGGAAGAGACAGATCTCTGACAGTTGCACTACT 1260
    |||
Db 1220 TCATAAGGAGAGTCAACCACTGTGTGGAAGAGACAGATCTCTGACAGTTGCACTACT 1279
OY 1261 TACCTTTGGCTAA 1273
    |||
Db 1280 TACCTTTGGCTAA 1292
```

```
RESULT 4
US-09-724-676A-44634
: Sequence 44634, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724, 676A
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 44634
: LENGTH: 2019
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1927)..(1927)
: OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-44634
```

```
Query Match          99.6%: Score 1268.2: DB 5: Length 2019:
Best Local Similarity 99.8%: Pred. No. 0:
Matches 1270: Conservative 0: Mismatches 3: Indels 0: Gaps 0:
```

```
QY 1 ATGCCGACAGCGGCGCCGAGCGGCGGCGTGCCTGTGACGCGCGCGCTCCCGAAGT 60
DB 20 ATGCCGACAGCGGCGCCGAGCGGCGGCGTGCCTGTGACGCGCGCGCTCCCGAAGT 79
QY 61 GTGCCGGGCGTGGCGGAAGTTTACAGAGGAGCGCGTGGCGCGGCGCGCGCTCCCGGC 120
DB 80 GTGCCGGGCGGCGGGAAGTTTACAGAGGAGCGCGTGGCGCGGCGCGCGCTCCCGGC 139
QY 121 ACGTGTCTGCGGACGTGGCAGCGCGCGCTGCCTGGCTTGGAGGCGCGCGCGCTTGA 180
DB 140 ACGTGTCTGCGGACGTGGCAGCGCGCGCTGCCTGGCTTGGAGGCGCGCGCGCTTGA 199
QY 181 TCGCGCGGCGGTGTGCGCGAGTGCGTGTGCTTTCACATCGCGGACGCGCTTATTC 240
DB 200 TCGCGCGGCGGTGTGCGCGAGTGCGTGTGCTTTCACATCGCGGACGCGCTTATTC 259
QY 241 ACGCGGCGAGAGCTCAGCTCGACCGCGCGCGCGCGCTTCCAGAGGTTGATCTG 300
DB 260 ACGCGGCGAGAGCTCAGCTCGACCGCGCGCGCGCGCTTCCAGAGGTTGATCTG 319
QY 301 CTCCTTCCCTACTGTCAGCGCGCAATTTACAGCAAAAGCGCTTGAAGCGCGCTGTG 360
DB 320 CTCCTTCCCTACTGTCAGCGCGCAATTTACAGCAAAAGCGCTTGAAGCGCGCTGTG 379
QY 361 CAAAGACACGGGGGAGAGACATTTGTTGTGACTATGAAGGTTGGGAAAGCGCTTAT 420
DB 380 CAAAGACACGGGGGAGAGACATTTGTTGTGACTATGAAGGTTGGGAAAGCGCTTAT 439
QY 421 CAGGAGCTACATGTGAGCCCGCACATTTGACTCACACAGGAAAGCGCTTGTGTTG 480
DB 440 CAGGAGCTACATGTGAGCCCGCACATTTGACTCACACAGGAAAGCGCTTGTGTTG 499
QY 481 TGCAGCCTGCTGGCTGTGATCAAAAATTCAACAGAAAATAAATTGAATAAATTTTGA 540
DB 500 TGCAGCCTGCTGGCTGTGATCAAAAATTCAACAGAAAATAAATTGAATAAATTTTGA 559
QY 541 ACGCAACATGAAATCAACAAAAAATATATATGAGTTTGAAGACTGTAGAGAGC 600
DB 560 ACGCAACATGAAATCAACAAAAAATATATATGAGTTTGAAGACTGTAGAGAGC 619
QY 601 CTTTAAAGAAATGAGAGCTGAAATTCATCATGTCGAGCATACCAATGAGACCTCTATT 660
DB 620 CTTTAAAGAAATGAGAGCTGAAATTCATCATGTCGAGCATACCAATGAGACCTCTATT 679
QY 661 CAAGTGTACCGAGGAAGATGTGGGAAACAATTGATCATACCGAGCAAGCTGAAAGACA 720
DB 680 CAAGTGTACCGAGGAAGATGTGGGAAACAATTGATCATACCGAGCAAGCTGAAAGACA 739
```

```
QY 721 TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 780
DB 740 TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 799
QY 781 GAGGAGACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATACTATGAGAGTATG 840
DB 800 GAGGAGACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATACTATGAGAGTATG 859
QY 841 CCGGAAACATTTAAACCCAAAGTTTACCTTAAGCAACATGAAACATCATGCCCGACA 900
DB 860 CCGGAAACATTTAAACCCAAAGTTTACCTTAAGCAACATGAAACATCATGCCCGACA 919
QY 901 AAGGATGTATGCTGCTGTCCAAAGAGGCTGTGAGAACCTACTACTGTGTTTAA 960
DB 920 AAGGATGTATGCTGCTGTCCAAAGAGGCTGTGAGAACCTACTACTGTGTTTAA 979
QY 961 TCTCCAAAGCCATATCCTCTCTTCATGAGGAAAGCCGCTTTTGTGTGAAACATGC 1020
DB 980 TCTCCAAAGCCATATCCTCTCTTCATGAGGAAAGCCGCTTTTGTGTGAAACATGC 1039
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTAGGATGCTGTGATCATGA 1080
DB 1040 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTAGGATGCTGTGATCATGA 1099
QY 1081 TCTGTACAGAAAGAAATGAGAGCTCAAGTCAAAAATCTCTGGA AAAACGAGTTTGGC 1140
DB 1100 TCTGTACAGAAAGAAATGAGAGCTCAAGTCAAAAATCTCTGGA AAAACGAGTTTGGC 1159
QY 1141 CTCCTATCTCAGTGTGATATCCCTCCCAAGGAAACAGGCAAGGCTTATCTTTGTG 1200
DB 1160 CTCCTATCTCAGTGTGATATCCCTCCCAAGGAAACAGGCAAGGCTTATCTTTGTG 1219
QY 1201 TCAAAAGGAGAGTACCCACATGTGTGAAAGCAAGATCTCTGACAGTTGCACTACT 1260
DB 1220 TCAAAAGGAGAGTACCCACATGTGTGAAAGCAAGATCTCTGACAGTTGCACTACT 1279
QY 1261 TACCTTGGCTTAA 1273
DB 1280 TACCTTGGCTTAA 1292
```

```
RESULT 5
US-09-724-676-44645
: Sequence 44645, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724, 676
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 44645
: LENGTH: 2104
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-724-676-44645
```

```
Query Match          99.6%: Score 1268.2: DB 5: Length 2104:
Best Local Similarity 99.8%: Pred. No. 0:
Matches 1270: Conservative 0: Mismatches 3: Indels 0: Gaps 0:
```

```
QY 1 ATGCCGACAGCGGCGCGGCGGCGTGCCTGTGACGCGCGCGCTCCCGAAGT 60
DB 20 ATGCCGACAGCGGCGCGGCGGCGGCGTGCCTGTGACGCGCGCGCTCCCGAAGT 79
QY 61 GTGCCGGGCGTGGCGGAAGTTTACAGAGGAGCGCGTGGCGCGGCGCGCTTCCCGGC 120
DB 80 GTGCCGGGCGGCGGGAAGTTTACAGAGGAGCGCGTGGCGCGGCGCGCTTCCCGGC 139
QY 121 ACGTGTCTGCGGACGTGGCAGCGCGCGCTGCCTGGCTTGGAGGCGCGCGCGCTTGA 180
DB 140 ACGTGTCTGCGGACGTGGCAGCGCGCGCTGCCTGGCTTGGAGGCGCGCGCGCTTGA 199
```

Db 140 AGGTGCTGGGACGTGGCAGCGGCGCTGGCCCTGGGCTTGGAGGCGCGCGCCCTGGA 199
QY 181 TCCGCCGCGCCGTGGTCCGCGAGTGGGTGTGCTCTTGACCATGCGCGACGCGCTCATTCG 240
Db 200 TCCGCCGCGCGCGGTGGTCCGCGAGTGGGTGTGCTCTTGACCATGCGCGACGCGCTCATTCG 259
QY 241 ACCCGGCGAGACGTGAGTCCGACCGCGCGCGCGCGCGCTTCCGAGAGGTTATGCG 300
Db 260 ACCCGGCGAGAGCTGAGTCCGACCGCGCGCGCGCGCGCGCTTCCGAGAGGTTATGCG 319
QY 301 CTCTCTCCCTGACTGACGCGCCCATTTACAGCAAAAGCTTGAGAGCTTGAGCGGACCTGTG 360
Db 320 CTCTCTCCCTGACTGACGCGCCCATTTACAGCAAAAGCTTGAGAGCTTGAGCGGACCTGTG 379
QY 361 CAAGGACAGGGGAGAGACCATTTGTTTGTGACTATGAGGGGTGGAGAGGCTTCAT 420
Db 380 CAAGGACAGGGGAGAGACCATTTGTTTGTGACTATGAGGGGTGGAGAGGCTTCAT 439
QY 421 CAGGGAATACATCTGAGCGCGCCACATTTGACTCAGACAGGAAAGCCGTTGTG 480
Db 440 CAGGGAATACATCTGAGCGCGCCACATTTGACTCAGACAGGAAAGCCGTTGTG 499
QY 481 TGCAGCACTGGCTGTGATCAAAAATTCACACAAATCAAACTTGAGAAACATTTTGA 540
Db 500 TGCAGCACTGGCTGTGATCAAAAATTCACACAAATCAAACTTGAGAAACATTTTGA 559
QY 541 AGGCAAAACATGAAATCAACAAAACATATATATGAGTTTGAAGACTGTAGAGAC 600
Db 560 AGGCAAAACATGAAATCAACAAAACATATATATGAGTTTGAAGACTGTAGAGAC 619
QY 601 CTTTAAAGAACATCAGAGCTGAAATTCATAGTGCAGCATACCAATGAACTCTATT 660
Db 620 CTTTAAAGAACATCAGAGCTGAAATTCATAGTGCAGCATACCAATGAACTCTATT 679
QY 661 CAAGGTATCCAGGAAGATGTGGAAACATTTGCATCCAGCAAGCTGAAGAGACA 720
Db 680 CAAGGTATCCAGGAAGATGTGGAAACATTTGCATCCAGCAAGCTGAAGAGACA 739
QY 721 TGCAGGCGCCAGAGGCTGTATGTGCAAAAAGATGTCCTTGTGGCAAAAACATG 780
Db 740 TGCAGGCGCCAGAGGCTGTATGTGCAAAAAGATGTCCTTGTGGCAAAAACATG 799
QY 781 GAGGAACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 840
Db 800 GAGGAACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 859
QY 841 CCGGAAAACATTTAAACGCAAGATTAACCTTAAGCAACATGAAACCATGCCCCAGA 900
Db 860 CCGGAAAACATTTAAACGCAAGATTAACCTTAAGCAACATGAAACCATGCCCCAGA 919
QY 901 AAGGAGTATGTCCCTCTCCAAAGAAAGCTGTGGAAGACTTACTACTGTGTTAA 960
Db 920 AAGGAGTATGTCCCTCTCCAAAGAAAGCTGTGGAAGACTTACTACTGTGTTAA 979
QY 961 TCTCCAAAGCCATATCTCTCTCTCCATGAGAAAGCCGCTTTTGTGTGTAACATG 1020
Db 980 TCTCCAAAGCCATATCTCTCTCTCCATGAGAAAGCCGCTTTTGTGTGTAACATG 1039
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGCTCTACAGCATGCTGTGTATGTA 1080
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGCTCTACAGCATGCTGTGTATGTA 1099
QY 1081 TCTGCAAGAAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGAAAAACGAGTTTGGC 1140
Db 1100 TCTGCAAGAAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGAAAAACGAGTTTGGC 1159
QY 1141 CTCTCATCTCACTGATATATCCCTCCCAAAAGAAACAGAGGCAAGGCTTATCTTTGG 1200
Db 1160 CTCTCATCTCACTGATATATCCCTCCCAAAAGAAACAGAGGCTTATCTTTGG 1219
QY 1201 TCMAAACGAGAGTACCCCACTGTGTGGAAGACAGATGCTCTGACAGTTTGCATGACT 1260
Db 1220 TCMAAACGAGAGTACCCCACTGTGTGGAAGACAGATGCTCTCTGACAGTTTGCATGACT 1279

QY 1261 TACCTTGGCTAA 1273
Db 1280 TACCTTGGCTAA 1292

RESULT 6
US-09-724-676A-44645
; Sequence 44645, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 44645
; LENGTH: 2104
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-44645

Query Match 99.6% Score 1268.2; DB 5; Length 2104;
Best local Similarity 99.8% Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCGAGCAGCGCGCGCGAGCGCGGCGGTGCTGTGACCGCGCGCTCCGGAAGT 60
Db 20 ATGCGAGCAGCGCGCGCGAGCGCGGCGGTGCTGTGACCGCGCGCTCCGGAAGT 79
QY 61 GTCGCGCGCGTGGCGGAAGGTTCAAGAGGAGCGCGTGGCGCGCGCGCGTTCGCGG 120
Db 80 GTCGCGCGCGCGCGGAAGGTTCAAGAGGAGCGCGTGGCGCGCGCGCGTTCGCGG 139
QY 121 ACCTGTCTGCGGACGTGGCAGCGCGCGCTGGGCGCTTGGAGGCGCGCGCGCTGGA 180
Db 140 ACCTGTCTGCGGACGTGGCAGCGCGCGCTGGGCGCTTGGAGGCGCGCGCGCTGGA 199
QY 181 TCCGCGCGCGCGTGGTGGCGGAGTGGTGTGCTCTTACCATGCGCGAGCGCTCATTCG 240
Db 200 TCCGCGCGCGCGTGGTGGCGGAGTGGTGTGCTCTTACCATGCGCGAGCGCTCATTCG 259
QY 241 AGCGGCGAGAGCTGAGCTCCGACCGCGCGCGCGCTTCCAGAGGTTCAATCG 300
Db 260 AGCGGCGAGAGCTGAGCTCCGACCGCGCGCGCGCGCTTCCAGAGGTTCAATCG 319
QY 301 CTCTCTCCCTGACTGACGCGCAATTAAGCAAGCTGGAAGCTGAGCGGACCTGTG 360
Db 320 CTCTCTCCCTGACTGACGCGCAATTAAGCAAGCTGGAAGCTGAGCGGACCTGTG 379
QY 361 CAAGCAACAGGGGAGAGACCATTTGTTGTGACTATGAAGGCTGTGGAAGCGCTTCAT 420
Db 380 CAAGCAACAGGGGAGAGACCATTTGTTGTGACTATGAAGGCTGTGGAAGCGCTTCAT 439
QY 421 CAGGGAATACATCTGAGCGCGCCACATTTGACTCAGACAGGAAAGCCGTTGTG 480
Db 440 CAGGGAATACATCTGAGCGCGCCACATTTGACTCAGACAGGAAAGCCGTTGTG 499
QY 481 TGCAGCACTGGCTGTGATCAAAAATTCACACAAATCAAACTTGAGAAACATTTTGA 540
Db 500 TGCAGCACTGGCTGTGATCAAAAATTCACACAAATCAAACTTGAGAAACATTTTGA 559
QY 541 AGGCAAAACATGAAATCAACAAAACATATATATGAGTTTGAAGAGCTGTAAAGAC 600
Db 560 AGGCAAAACATGAAATCAACAAAACATATATATGAGTTTGAAGAGCTGTAAAGAC 619
QY 601 CTTTAAAGAACATCAGAGCTGAAATTCATAGTGCAGCATACCAATGAACTCTATT 660
Db 620 CTTTAAAGAACATCAGAGCTGAAATTCATAGTGCAGCATACCAATGAACTCTATT 679
QY 661 CAAGGTATCCAGGAAGATGTGGAAACATTTGCATCCAGCAAGCTGAAGAGACA 720

Db 680 CAAGTGTACCCGAGAGATGTGGGAAACATTTCATCCACGACACTGAAACGACA 739
Qy 721 TCGCAAGGCCCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 780
Db 740 TCCCAAGGCCCCACGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 799
Qy 781 GACGGAACCTTGAACATGTGAGAGAAACCATTAAGAGCAATCTATGTGAAGATG 840
Db 800 GACGGAACCTTGAACATGTGAGAGAAACCATTAAGAGCAATCTATGTGAAGATG 859
Qy 841 CCGGAAACATTTAAAGCCAAACATTCCTTAAGCAACATGAAACATCATGCCGACA 900
Db 860 CCGGAAACATTTAAAGCCAAACATTCCTTAAGCAACATGAAACATCATGCCGACA 919
Qy 901 AAGCATGTATGTCCGTGTCCAAAGAGAGCTGTGAGAACTATACTATGTGTTAA 960
Db 920 AAGCATGTATGTCCGTGTCCAAAGAGAGCTGTGAGAACTATACTATGTGTTAA 979
Qy 961 TCTCCAAAGCCATATCTCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGC 1020
Db 980 TCTCCAAAGCCATATCTCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGC 1039
Qy 1021 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAGAGCATGTCTGTACATGA 1080
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAGAGCATGTCTGTACATGA 1099
Qy 1081 TCTGACAAAGAAATGAAAGTCAAAAGTCAAAATCTCTGTGAAAACGAGTTGGC 1140
Db 1100 TCTGACAAAGAAATGAAAGTCAAAAGTCAAAATCTCTGTGAAAACGAGTTGGC 1159
Qy 1141 CTCTCATCTCACTGATATATCCCTCCCAAGAGCAAGGCAAGCTTATCTTTG 1200
Db 1160 CTCTCATCTCACTGATATATCCCTCCCAAGAGCAAGGCAAGCTTATCTTTG 1219
Qy 1201 TCMAAAGGAGAGTACCCACATGTGTGAAGAGACAGATCTCTGACAGTTGCACT 1260
Db 1220 TCMAAAGGAGAGTACCCACATGTGTGAAGAGACAGATCTCTGACAGTTGCACT 1279
Qy 1261 TACCCCTGGCTAA 1273
Db 1280 TACCCCTGGCTAA 1292

RESULT 7
US-09-724-676-44678
Sequence 44678, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
FILE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724, 676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 44678
LENGTH: 2239
TYPE: DNA
ORGANISM: Homo sapiens
US-09-724-676-44678

Query Match 99.6%; Score 1268.2; DB 5; Length 2239;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGCCAGACAGCGCGCGGAGGCTGTGACCGCGGCGCTCCCGAAGT 60
Db 20 ATGCCAGACAGCGCGCGGAGGCTGTGACCGCGGCGCTCCCGAAGT 79
Qy 61 GTGCGGCGCTCGCGGAAGTTTCAGCAGGAGCCGCTGGCGCGCGCTTCGCGC 120
Db 80 GTGCGGCGCGCTCGCGGAAGTTTCAGCAGGAGCCGCTGGCGCGCGCTTCGCGC 139
Qy 121 ACGTGTCTGCGACGCTGCGACGCGGCTGGCGCTGGCGCTGGAGCGCGCGCTTGA 180

Db 140 ACGTGTCTGCGACGCTGCGACGCGGCTGGCGCTGGCGCTGGAGCGCGCGCTTGA 199
Qy 181 TCGCGCGCGCTGCTGCGCGAGTGTGCTCTTCCTTGTACCATGCGGAGCGCTTATTC 240
Db 200 TCGCGCGCGCTGCTGCGCGAGTGTGCTCTTCCTTGTACCATGCGGAGCGCTTATTC 259
Qy 241 ACGCGCGGAGCTAGGCTCGGACCGCGCGCGCGCGCTTCCTCCGAGAGCTTCATCG 300
Db 260 ACGCGCGGAGCTAGGCTCGGACCGCGCGCGCGCGCTTCCTCCGAGAGCTTCATCG 319
Qy 301 CTCTCCCTGACTGCTGACGCGCAATTCAGCAAGCCCTGAGACCTTGACGCGACTGTG 360
Db 320 CTCTCCCTGACTGCTGACGCGCAATTCAGCAAGCCCTGAGACCTTGACGCGACTGTG 379
Qy 361 CAAGCAGCGGCGGAGACCATTTGTTGTGACTATGAAGGCTGTGAGAGCGCTTAT 420
Db 380 CAAGCAGCGGCGGAGACCATTTGTTGTGACTATGAAGGCTGTGAGAGCGCTTAT 439
Qy 421 CAGGACTACCATCTGACCGCGCAATTCGACTCAGCAGAGAGAAACCGCTTTGTTG 480
Db 440 CAGGACTACCATCTGACCGCGCAATTCGACTCAGCAGAGAGAAACCGCTTTGTTG 499
Qy 481 TCGAGCCACTGCTGTGATCAAAAATTCACACAGAAATCAAACTTGAAGAAACATTTGA 540
Db 500 TCGAGCCACTGCTGTGATCAAAAATTCACACAGAAATCAAACTTGAAGAAACATTTGA 559
Qy 541 ACCCAACATGAATAATCAACAAAAATATATATGACATTTTGAAGACTGTAAAGAC 600
Db 560 ACCCAACATGAATAATCAACAAAAATATATATGACATTTTGAAGACTGTAAAGAC 619
Qy 601 CTTTAAAGAACTGACGAGCTGAATTCATAGTGCACATACCAATGAACCTCTAT 660
Db 620 CTTTAAAGAACTGACGAGCTGAATTCATAGTGCACATACCAATGAACCTCTAT 679
Qy 661 CAAGTGTACCCAGAGAGATGTGGAAACATTTGCAATCCAGCAAGCTGAAACGACA 720
Db 680 CAAGTGTACCCAGAGAGATGTGGAAACATTTGCAATCCAGCAAGCTGAAACGACA 739
Qy 721 TGGCAAGGCCCCAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 780
Db 740 TGGCAAGGCCCCAGAGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACATG 799
Qy 781 GACGGAACCTTGAACATGTGAGAGAAACCATTAAGAGCAATCTATGTGAAGATG 840
Db 800 GACGGAACCTTGAACATGTGAGAGAAACCATTAAGAGCAATCTATGTGAAGATG 859
Qy 841 CCGGAAACATTTAAAGCCAAACATTCCTTAAGCAACATGAAACATCATGCCGACA 900
Db 860 CCGGAAACATTTAAAGCCAAACATTCCTTAAGCAACATGAAACATCATGCCGACA 919
Qy 901 AAGGATGTATGTCTCTGTCCAGAGAGGCTGTGAAGAACCTTACTACTGTGTTAA 960
Db 920 AAGGATGTATGTCTCTGTCCAGAGAGGCTGTGAAGAACCTTACTACTGTGTTAA 979
Qy 961 TCTCCAAAGCCATATCTCTCTTCATGAGAGAGCGCGCTTTGTGTGAACATGC 1020
Db 980 TCTCCAAAGCCATATCTCTCTTCATGAGAGAGCGCGCTTTGTGTGAACATGC 1039
Qy 1021 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAGAGCATGTCTGTACATGA 1080
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAGAGCATGTCTGTACATGA 1099
Qy 1081 TCTGACAAAGAAATGAAAGTCAAAAGTCAAAATCTCTGTGAAAACGAGTTTGGC 1140
Db 1100 TCTGACAAAGAAATGAAAGTCAAAAGTCAAAATCTCTGTGAAAACGAGTTTGGC 1159
Qy 1141 CTCTCATCTCACTGATATATCCCTCCCAAGAGCAAGGCAAGCTTATCTTTG 1200
Db 1160 CTCTCATCTCACTGATATATCCCTCCCAAGAGCAAGGCAAGCTTATCTTTG 1219
Qy 1201 TCMAAAGGAGAGTACCCACATGTGTGAAGAGACAGATCTCTGACAGTTGCACT 1260


```
Db 1220 TCAGAACGGAGAGTACCCACACTGTGTGGAAGAACATGCTCTGACAGTTGCAGTACT 1279
QY 1261 TACCTTGGCTTAA 1273
    |||||
Db 1280 TACCTTGGCTTAA 1292

RESULT 8
US-09-724-676A-44678
; Sequence 44678, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 44678
; LENGTH: 2239
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-44678

Query Match          99.8%; Score 1268.2; DB 5; Length 2239;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCGCAGACAGCGCGCCGACGCGGGCGGTGCTGTGACCGCGCGCTCCCGAAGT 60
    |||||
Db 20 ATGCGCAGAGGCGCGCCGACGCGGGCGGTGCTGTGACCGCGCGCTCCCGAAGT 79
    |||||

QY 61 GTGCGCGCGGTGCGCGGAGGTTTCAAGAGGAGCGGTGGCGCGGGCGGTGCTCCGCG 120
    |||||
Db 80 GTGCGCGCGCGCGGAGAGGTTTCAAGAGGAGCGGTGGCGCGGGCGGTGCTCCGCG 139
    |||||

QY 121 ACGTGTCTGCGCAGCTGCGACGCGCGCTTGCGCTTGAGAGCGCGCGCGCTTGA 180
    |||||
Db 140 ACGTGTCTGCGCAGCTGCGACGCGCGCTTGCGCTTGAGAGCGCGCGCGCTTGA 199
    |||||

QY 181 TCGCGCGCGCGGTGCGCGGAGTGGTGTGCTGCTTGAACCATGGCGAGCGCTTATGTC 240
    |||||
Db 200 TCGCGCGCGCGGTGCGCGGAGTGGTGTGCTGCTTGAACCATGGCGAGCGCTTATGTC 259
    |||||

QY 241 AGCGGCGGAGAGCTACGCTCCGACCCGCGCGCGCGCTTCCAGAGGTTGATCTG 300
    |||||
Db 260 AGCGGCGGAGAGCTACGCTCCGACCCGCGCGCGCGCTTCCAGAGGTTGATCTG 319
    |||||

QY 301 CTCCTTCCCTGACCTGACGCGCGCAATTACAGAAAGCTTGAAGCTTGACGACCTGTG 360
    |||||
Db 320 CTCCTTCCCTGACCTGACGCGCGCAATTACAGAAAGCTTGAAGCTTGACGACCTGTG 379
    |||||

QY 361 CAAAGCACAGGGGGGAGAGACCATTTTGTGACTATGAAGGTGTGGCAAGGCTTCAAT 420
    |||||
Db 380 CAAAGCACAGGGGGGAGAGACCATTTTGTGACTATGAAGGTGTGGCAAGGCTTCAAT 439
    |||||

QY 421 CAGGAGCTACATCTGAGCGCGCAATTTGACTCACAGAGGAAAGACCGCTTTGTTTG 480
    |||||
Db 440 CAGGAGCTACATCTGAGCGCGCAATTTGACTCACAGAGGAAAGACCGCTTTGTTTG 499
    |||||

QY 481 TGCAGGCACTGCGCTGTGATCAAAAAATTCACAAAAATCAAACTTGAAGAAACATTTTGA 540
    |||||
Db 500 TGCAGGCAATGCGCTGTGATCAAAAAATTCACAAAAATCAAACTTGAAGAAACATTTTGA 559
    |||||

QY 541 ACGCAAAACATGAAATTCACAAAAATATATATGAGTTTGAAGACGTGAAGAGAC 600
    |||||
Db 560 ACGCAAAACATGAAATTCACAAAAATATATATGAGTTTGAAGACGTGAAGAGAC 619
    |||||

QY 601 CTTTAAAGAAACATGACGCGTGAATTCATGAGTGCACGACATACCAATGAACCTCTATT 660
    |||||
Db 620 CTTTAAAGAAACATGACGCGTGAATTCATGAGTGCACGACATACCAATGAACCTCTATT 679
    |||||

QY 661 CAAGGTACCCAGGAAGATGTGGAAACACTTTGATCAACCCAGCAAGCTGAAGAGACA 720
```

```
Db 680 CAAGGTACCCAGGAAGATGTGGAAACACTTTGATCAACCCAGCAAGCTGAAGAGACA 739
    |||||
QY 721 TCGCAAGGCCCAAGAGGCTATGTATGTCAAAAAAGATGTCCCTTTGTGGCAAAACATG 780
    |||||
Db 740 TCGCAAGGCCCAAGAGGCTATGTATGTCAAAAAAGATGTCCCTTTGTGGCAAAACATG 799
    |||||

QY 781 GACGAGCTTGTGAACATGTGTGAGAAACCATTAAGAGGAATATATGATGGAAGTATG 840
    |||||
Db 800 GACGAGCTTGTGAACATGTGTGAGAAACCATTAAGAGGAATATATGATGGAAGTATG 859
    |||||

QY 841 CCGGAAACATTTAAACGCAAAAGATTACTTTAAGCAACATGAAACATCATGCCCCAGA 900
    |||||
Db 860 CCGGAAACATTTAAACGCAAAAGATTACTTTAAGCAACATGAAACATCATGCCCCAGA 919
    |||||

QY 901 AAGGATGTATGTGCTGTGCTCCAGAGAAAGCTGTGGAAGACCTATGACTGTGTTTAA 960
    |||||
Db 920 AAGGATGTATGTGCTGTGCTCCAGAGAAAGCTGTGGAAGACCTATGACTGTGTTTAA 979
    |||||

QY 961 TCTCCAAAGCATATCTCTCTCCATGAGGAAAGCGCCCTTTGTGTGGAACATGC 1020
    |||||
Db 980 TCTCCAAAGCATATCTCTCTCCATGAGGAAAGCGCCCTTTGTGTGGAACATGC 1039
    |||||

QY 1021 TGGCTGTGCAAAACATTTGCAATGAACAAAGCTGTGACTATGCTGTGTCACATGA 1080
    |||||
Db 1040 TGGCTGTGCAAAACATTTGCAATGAACAAAGCTGTGACTATGCTGTGTCACATGA 1099
    |||||

QY 1081 TCTGTCAAGAAAGAAATGAAGCTCAAAAGTCAAAAAATCTCTGAAAAACGAGTTTGGC 1140
    |||||
Db 1100 TCTGTCAAGAAAGAAATGAAGCTCAAAAGTCAAAAAATCTCTGAAAAACGAGTTTGGC 1159
    |||||

QY 1141 CTCCTATCTCACTGATGATATCTCTCCCAAGAAACAAAGGCAAGGCTTATCTTTGTG 1200
    |||||
Db 1160 CTCCTATCTCACTGATGATATCTCTCCCAAGAAACAAAGGCAAGGCTTATCTTTGTG 1219
    |||||

QY 1201 TCAGAACGGAGAGTACCCACACTGTGTGGAAGAACATGCTTCGACAGTTGCACTACT 1260
    |||||
Db 1220 TCAGAACGGAGAGTACCCACACTGTGTGGAAGAACATGCTTCGACAGTTGCACTACT 1279
    |||||

QY 1261 TACCTTGGCTTAA 1273
    |||||
Db 1280 TACCTTGGCTTAA 1292
    |||||

RESULT 9
US-09-724-676-44656
; Sequence 44656, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 44656
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-44656

Query Match          99.8%; Score 1268.2; DB 5; Length 2271;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCGCAGCAGCGCGCGCGGCGCGGTGCTGTGACCGCGCGCTCCCGAAGT 60
    |||||
Db 20 ATGCGCAGCAGCGCGCGCGGCGCGGTGCTGTGACCGCGCGCTCCCGAAGT 79
    |||||

QY 61 GTGCGCGCGCTTCCCGGAGGTTTCAAGAGGAGCGCGGTGGCGCGGGCGGTGCTCCGCG 120
    |||||
Db 80 GTGCGCGCGCGCGGAGGTTTCAAGAGGAGCGCGGTGGCGCGGGCGGTGCTCCGCG 139
    |||||
```

```
QY 121 ACGTCTCGACCTGGCAGCGCCCTGGGCTTGAGGCGCGCCCTGGA 180
    |||
Db 140 ACGTGTCTGGGACCTGGCAGCGCCCTGGGCTTGAGGCGCGCCCTGGA 199
QY 181 TCCGCGGCGCGGTGTGCGCGAGTCGCTGCTTGGACCATGCGCGCGCTTATTC 240
    |||
Db 200 TCCGCGGCGCGGTGTGCGCGAGTCGCTGCTTGGACCATGCGCGCGCTTATTC 259
QY 241 AGCGGCGGAGCTGAGCTCGGACCGCGCGCGCGCGCGCTTCCGAGAGTTCACTG 300
    |||
Db 260 AGCGGCGGAGCTGAGCTCGGACCGCGCGCGCGCGCGCTTCCGAGAGTTCACTG 319
QY 301 CTCCTCCCTGACTGACGCGGCAATTACAGAAAGCTTGAAGCGCGACCTGTG 360
    |||
Db 320 CTCCTCCCTGACTGACGCGGCAATTACAGAAAGCTTGAAGCGCGACCTGTG 379
QY 361 CAAGCACAGCGGCGAGACATTTGTTGATGATGAGAGGCTGCGCAAGCGCTTAT 420
    |||
Db 380 CAAGCACAGCGGCGAGACATTTGTTGATGATGAGAGGCTGCGCAAGCGCTTAT 439
QY 421 CAGGACTACCATCTGAGCGCGCACTTGTGACTCACAGAGAAAGCGCTTTGTTG 480
    |||
Db 440 CAGGACTACCATCTGAGCGCGCACTTGTGACTCACAGAGAAAGCGCTTTGTTG 499
QY 481 TGGAGCACTGGCTGTGATCAAAAATTCACACAAATTCGAAGTTGAAGAAACATTTTGA 540
    |||
Db 500 TGGAGCACTGGCTGTGATCAAAAATTCACACAAATTCGAAGTTGAAGAAACATTTTGA 559
QY 541 AGCCAAACATGAAATTCACAAACAAATATATATGACGTTTGAAGACTGTGAAGAGC 600
    |||
Db 560 AGCCAAACATGAAATTCACAAACAAATATATATGACGTTTGAAGACTGTGAAGAGC 619
QY 601 CTTTAAAGAACATCAGCAGCTGAAATTCATCATGAGTGCACATACAAAGAACCTCTATT 660
    |||
Db 620 CTTTAAAGAACATCAGCAGCTGAAATTCATCATGAGTGCACATACAAAGAACCTCTATT 679
QY 661 CAAGTGTACCGAGAGAGATGTGGAAACACTTGTGATCCAGCAGAGCTGAAGAGACA 720
    |||
Db 680 CAAGTGTACCGAGAGAGATGTGGAAACACTTGTGATCCAGCAGAGCTGAAGAGACA 739
QY 721 TCCCAAGGCGCCAGAGGCTATGTATGTCAAAAGAGATGTCCTTGTGGCAAAAACATG 780
    |||
Db 740 TCCCAAGGCGCCAGAGGCTATGTATGTCAAAAGAGATGTCCTTGTGGCAAAAACATG 799
QY 781 GACGCACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGATATG 840
    |||
Db 800 GACGCACTTCTGAAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGATATG 859
QY 841 CCGGAAACATTTAAACGCAAGATTAACCTTAAGCAACATGAAAGCATGAGCCGACGA 900
    |||
Db 860 CCGGAAACATTTAAACGCAAGATTAACCTTAAGCAACATGAAAGCATGAGCCGACGA 919
QY 901 AAGGAGTATATCGCTCTCCAGAGAGAGCTGTGAGAAAGCTTATGACTGTGTTAA 960
    |||
Db 920 AAGGAGTATATCGCTCTCCAGAGAGAGCTGTGAGAAAGCTTATGACTGTGTTAA 979
QY 961 TCTCCAAAGCCATATCTCTCTTCATGAGAGAAAGCGCGCTTTGTGTGTAAGCATTC 1020
    |||
Db 980 TCTCCAAAGCCATATCTCTCTTCATGAGAGAAAGCGCGCTTTGTGTGTAAGCATTC 1039
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTGACTAGGCATGCTGTGTACATGA 1080
    |||
Db 1040 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTGACTAGGCATGCTGTGTACATGA 1099
QY 1081 TCCCTGCAAGAGAAATGAGAGCTCAAAAGTCAAAAATTCGTCGAAAAACGAGTTTGGC 1140
    |||
Db 1100 TCCCTGCAAGAGAAATGAGAGCTCAAAAGTCAAAAATTCGTCGAAAAACGAGTTTGGC 1159
QY 1141 CTCCTATCTGAGTGGATATATCCCTCCCAAGAGAAACAGGCAAGGCTTATCTTTGTG 1200
    |||
Db 1160 CTCCTATCTGAGTGGATATATCCCTCCCAAGAGAAACAGGCAAGGCTTATCTTTGTG 1219
QY 1201 TCAGAAACGAGAGTACCCAACTGTGTGAAAGCAAGATGCTCTGACAGTTGACGATCT 1260
```

```
Db 1220 TCAGAAACGAGAGTACCCAACTGTGTGAGAGCAAGATGCTCTGACAGTTGACGATCT 1279
    |||
QY 1261 TACCCCTTGGCTTAA 1273
    |||
Db 1280 TACCCCTTGGCTTAA 1292
    |||

RESULT 10
US-09-724-676A-44656
: Sequence 44656, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 44656
: LENGTH: 2271
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-724-676A-44656

Query Match 99.68; Score 1268.2; DB 5; Length 2271;
Best Local Similarity 99.88; Pred. No. 0;
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCGAGAGCGGCGCGGACGCGGCGCGCTGTGTCACCGCGCGCTCCGGAAGT 60
    |||
Db 20 ATGCGAGAGCGGCGCGGACGCGGCGCGCTGTGTCACCGCGCGCTCCGGAAGT 79
QY 61 GTGCGGCGCTGCGCGGAAGTTCAAGAGAGCGGTGCGGCGCGCGCGCTTCCCGGC 120
    |||
Db 80 GTGCGGCGCGCGCGGAAGTTCAAGAGAGCGGTGCGGCGCGCGCGCTTCCCGGC 139
QY 121 ACGTGTCTGCGCAGCTGCGCAGCGCGCGCGCTGTGCTGTGAGAGCGCGCGCTTGA 180
    |||
Db 140 ACGTGTCTGCGCAGCTGCGCAGCGCGCGCGCTGTGCTGTGAGAGCGCGCGCTTGA 199
QY 181 TCCGCGGCGCGGTGTGCGCGAGTCGCTGCTTGGACATGCGCGAGCGCTTATTC 240
    |||
Db 200 TCCGCGGCGCGGTGTGCGCGAGTCGCTGCTTGGACATGCGCGAGCGCTTATTC 259
QY 241 AGCGGCGGAGCTTCACTTCCGACCCCGCGCGCGCGCGCTTCCAGAGAGTTGATCTG 300
    |||
Db 260 AGCGGCGGAGCTTCACTTCCGACCCCGCGCGCGCGCGCTTCCAGAGAGTTGATCTG 319
QY 301 CTCCTCCCTGACTGACGCGGCAATTACAGAAAGCTTGAAGCGCGACCTGTG 360
    |||
Db 320 CTCCTCCCTGACTGACGCGGCAATTACAGAAAGCTTGAAGCGCGACCTGTG 379
QY 361 CAAGCACAGCGGCGAGAGACATTTGTTGATGATGAGAGGCTGTGCAAGGCTTAT 420
    |||
Db 380 CAAGCACAGCGGCGAGAGACATTTGTTGATGATGAGAGGCTGTGCAAGGCTTAT 439
QY 421 CAGGACATCACTGAGCGCGGCACTTGTGACTGACAGAGAGAAAGCGCTTTGTTG 480
    |||
Db 440 CAGGACATCACTGAGCGCGGCACTTGTGACTGACAGAGAGAAAGCGCTTTGTTG 499
QY 481 TCGAGCACTGGCTGTGATCAAAAATTCACACAAATTCGAAGTTGAAGAAACATTTTGA 540
    |||
Db 500 TCGAGCACTGGCTGTGATCAAAAATTCACACAAATTCGAAGTTGAAGAAACATTTTGA 559
QY 541 AGCGAAACATGAAATTCAGAAACAAATATATATGAGTTTGAAGACTGTGAAGAGC 600
    |||
Db 560 AGCGAAACATGAAATTCAGAAACAAATATATATGAGTTTGAAGACTGTGAAGAGC 619
QY 601 CTTTAAAGAACATCAGCAGCTGAAATTCATGAGTGCACGATACCAATGAACCTCTATT 660
    |||
Db 620 CTTTAAAGAACATCAGCAGCTGAAATTCATGAGTGCACGATACCAATGAACCTCTATT 679
```

Yy 661 CAACTGACCCAGGAAGCATGTGGCAACACTTTGCATCACCAGCAAGCTGAACGACA 720
 ||| |||||
 Db 680 CAAGTGTACCCAGGAAGCATGTGGCAACACTTTGCATCACCAGCAAGCTGAACGACA 739
 Yy 721 TCGCAAGGCCCAAGAGGCTATGTATGTCCAAGAAAGAATGTTCTTTGTGGCAAAAACATG 780
 ||| |||||
 Db 740 TTGCCAAGGCCCAAGAGGCTATGTATGTCCAAGAAAGAATGTTCTTTGTGGCAAAAACATG 799
 Yy 781 GAGGGAACCTTGAAACATGTGASGAAGAACCCATTAAAGAGAAATACTATGTGAAGTAG 840
 ||| |||||
 Db 800 GAGGGAACCTTGAAACATGTGASGAAGAACCCATTAAAGAGAAATACTATGTGAAGTAG 859
 Yy 841 CCAGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAAATCTCATGCCCCAGA 900
 ||| |||||
 Db 860 CCAGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAAATCTCATGCCCCAGA 919
 Yy 901 AAGGATGTATGTGCGCTGTCCAAGAGAGCGTGTGSAAGAACCTATCTACTGTGTTTAA 960
 ||| |||||
 Db 920 AAGGATGTATGTGCGCTGTCCAAGAGAGCGTGTGSAAGAACCTATCTACTGTGTTTAA 979
 Yy 961 TTCACCAAGCATATCTCTCTTCCTTCATGASGAAGAGCGCCCTTTTGTGTGTAACATG 1020
 ||| |||||
 Db 980 TTCACCAAGCATATCTCTCTTCCTTCATGASGAAGAGCGCCCTTTTGTGTGTAACATG 1039
 Yy 1021 TGCGCTGTGCAAAACATTTGCAANTGAACAAAGTCTCACTAGGATCTTTGTACATGA 1080
 ||| |||||
 Db 1040 TGCGCTGTGCAAAACATTTGCAANTGAACAAAGTCTCACTAGGATCTTTGTACATGA 1099
 Yy 1081 TCCTGCACAGAAATAAATAAGCTCAAAGTCAAAAAATCTCTGTGAAAAAGGAGTTGGC 1140
 ||| |||||
 Db 1100 TCCTGCACAGAAATAAATAAGCTCAAAGTCAAAAAATCTCTGTGAAAAAGGAGTTGGC 1159
 Yy 1141 CTCCTCATCTCAGTGTATATNTCCCCTCCCAAAAGCAACAGGCGCTTATCTTTTGTG 1200
 ||| |||||
 Db 1160 CTCCTCATCTCAGTGTATATNTCCCCTCCCAAAAGCAACAGGCGCTTATCTTTTGTG 1219
 Yy 1201 TCAAAACGGAGATCACCCCAACGTGTGGAAGACAAAGTGTCTCTGSCACATCTGACATCT 1260
 ||| |||||
 Db 1220 TCAAAACGGAGATCACCCCAACGTGTGTGAAGACAAAGTGTCTCTGSCACATCTGACATCT 1279
 Yy 1261 TACCCTTGCGCTAA 1273
 ||| |||||
 Db 1280 TACCCTTGCGCTAA 1292
 RESULT 11
 US-09-724-676-44623
 : Sequence 44623; Application US/09724676
 : GENERAL INFORMATION:
 : APPLICANT: CompuGen LTD
 : TITLE OF INVENTION: Variants of alternative splicing
 : FILE REFERENCE: 129181.4 CompuGen
 : CURRENT APPLICATION NUMBER: US/09/724.676
 : NUMBER OF SEQ ID NOS: 97222
 : SOFTWARE: PatentIn version 3.2
 : SEQ ID NO 44623
 : LENGTH: 2352
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : US-09-724-676-44623

Query Match	99.68%	Score 1268.2	DB 5	Length 2352
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1270	Conservative 0	Mismatches 3	Indels 0	Gaps 0
QY	1	ATGCGCAGCAGCGCGGCCACCGCGGGCGCGTCTGTGACCGCGCGCGCTCCCGGAGT	60	
Db	20	ATGCGCAGCAGCGCGGCCACCGCGGGCGCGTCTGTGACCGCGCGCGCTCCCGGAGT	79	
QY	61	GTGCGCGCGTCTGCGCGCAGAGGTTTCAGCAGAGGAGCCGTGTGGGCCGGGCGCGCGGATTCCCGG	120	
Db	80	GTGCGCGCGCGCGCGCAGAGGTTTCAGCAGAGGAGCCGTGTGGGCCGGGCGCGCGGATTCCCGG	139	

Oy	121	ATGTGCTGAGCAGCTGGACAGCGACGCTGGCCCTGGGCGTTGGAGGCGCGCGCCCTGGGA	180
Db	140	ACGTGCTGGGACAGTGGCAGCGGCGCCCTGGCCCTTGGAGGCGCGCGCGCCCTGGGA	199
Oy	181	TGCGCCGGCGCGTGGTGGCGCGAGTGGTGTGTCTCTTTGGACATTCGGCGGACGGCTTCATTCG	240
Db	200	TTCGGCGGCGCGTGGTGGCGCGAGTGGTGTGTCTCTTGGACATTCGGCGGACGGCTTCATTCG	259
Oy	241	AGCGGCGAGAGCTCAGTCTCGACCGCGCGCGCGCGCGTCTCCAGAGAGCTTCATCTG	300
Db	260	AGCGGCGAGAGCTCAGCTCCGACCCCGCGCGCGCGCGCTTCCAGAGAGTTCATCTG	319
Oy	301	CTCCCTTCCCTGACCTGCAGCGCAATTAACAGCAAAAGCTGGAGCTTGCAGCGCACCTGTG	360
Db	320	CTCCCTTCCCTGACCTGCAGCGCAATTAACAGCAAAAGCTGGAGCTTGCAGCGCACCTGTG	379
Oy	361	CAAGGACAGGGGGGAGAGCAATTTGTGTGACTTGAAGGCTGTGGCAAGGCGCTTCAT	420
Db	380	CAAGGACAGGGGGGAGAGCAATTTGTGTGACTTGAAGGCTGTGGCAAGGCGCTTCAT	439
Oy	421	CAGGAGCATCATCTGAGCGCGCACATTTGTACTCAACAGAGAAAGCCGTTTGTTCG	480
Db	440	CAGGAGCATCATCTGAGCGCGCACATTTGTACTCAACAGAGAAAGCCGTTTGTTCG	499
Oy	481	TGCAGCCACGCGGTGGATCAAAAATTCACACAAAATCAAACTTGAAGAAACATTTTGA	540
Db	500	TGCAGCCCAATGGCTGTGATCAAAAATTCACACAAAATCAAACTTGAAGAAACATTTTGA	559
Oy	541	ACGCAAAACATGAAAAATCAACAAAAACAATATATATGCAGTTTGAAGACTGTAGAAGAC	600
Db	560	ACGCAAAACATGAAAAATCAACAAAAACAATATATATGCAGTTTGAAGACTGTAGAAGAC	619
Oy	601	CTTTAAAGAAACATCAGCAGCTGAAAAATCCATAGTCCAGACATPACCAATGAAACCTCATTT	660
Db	620	CTTTAAAGAAACATCAGCAGCTGAAAAATCCATAGTCCAGACATPACCAATGAAACCTCATTT	679
Oy	661	CAAGTGCATCCACAGGAAGATGTGGGAAACACTTTGCATCACCCAGCAAGCTTGAAGAGCA	720
Db	680	CAAGTGCATCCACAGGAAGATGTGGGAAACACTTTGCATCACCCAGCAAGCTTGAAGAGCA	739
Oy	721	TGCGAAGGCGCCAGGAGGCTTGTGATGTCAAAAAAGGACTCTCCCTTTGAGGCAAAACATG	780
Db	740	TGCGAAGGCGCCAGGAGGCTTGTGATGTCAAAAAAGGACTCTCCCTTTGAGGCAAAACATG	799
Oy	781	GAGGGAACTTTCTGAAACATGTGAGAGAAACCCATTAAGAGGAATACTATGTGAAGTATG	840
Db	800	GAGGGAACTTTCTGAAACATGTGAGAGAAACCCATTAAGAGGAATACTATGTGAAGTATG	859
Oy	841	CCGGAAGAACTTTAAACGAAAGATTCCTTAAAGCAACCATGAAACATCATGGCCGACGA	900
Db	860	CCGGAAGAACTTTAAACGCAAGATTAACCTTAAAGCAACCATGAAACATCATGGCCGACGA	919
Oy	901	AAGGATGTATGTCCGTCTCAAGAGAGGCGTGTGAAGAACCTTACTACTGTGTTTAA	960
Db	920	AAGGATGTATGTCCGTCTCAAGAGAGGCGTGTGAAGAACCTTACTACTGTGTTTAA	979
Oy	961	TCTGCAAGGCATATCTCTCTCTTCATAGAGAAAGCGCGCTTTTGTGTGAACATGC	1020
Db	980	TCTGCAAGGCATATCTCTCTCTTCATAGAGAAAGCGCGCTTTTGTGTGAACATGC	1039
Oy	1021	TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCACTAGGCAATGCTGTTGTACATGA	1080
Db	1040	TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCACTAGGCAATGCTGTTGTACATGA	1099
Oy	1081	TCTGCAAGAGAAATAATGAAGCTCAAAAGTCAAAAATCTCTGTAAAAAAGCAATTTGGC	1140
Db	1100	TCTGCAAGAGAAATAATGAAGCTCAAAAGTCAAAAATCTCTGTAAAAAAGCAATTTGGC	1159
Oy	1141	CTCTCATCTCAGTGGATATATCTCTCCCAAAAGAAACAAGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGGATATATCTCTCCCAAAAGAAACAAGGCAAGGCTTATCTTTGTG	1219

Db 80 GTGCCGCGCCGCGGAGGTTCAAGAGGAGCCGTGGCGCGCGCGGTTCCCGGC 139
 QY 121 AGGTGTCGTGGACGTGGCAGCGCGCTGGCCCTGGGCTTGGAGCGCGCGCGCTTGA 180
 Db 140 AGGTGTCGTGGACGTGGCAGCGCGCTGGCCCTGGGCTTGGAGCGCGCGCGCTTGA 199
 QY 181 TCCGCGCGCGCGGTGTGGCGGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 240
 Db 200 TCCGCGCGCGCGGTGTGGCGGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 259
 QY 241 ACCCGCGGAGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 300
 Db 260 ACCCGCGGAGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 319
 QY 301 CTCCTTCCCTGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 360
 Db 320 CTCCTTCCCTGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 379
 QY 361 CAAGCAGCGCGCGGAGACATTTTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 420
 Db 380 CAAGCAGCGCGCGGAGACATTTTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 439
 QY 421 CAGGAGTACATGTGGAGCGCGCACATTTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 480
 Db 440 CAGGAGTACATGTGGAGCGCGCACATTTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 499
 QY 481 TGCAGCTACTGCTGTGTGTCAAAAATTCACACAAAATCAAACTTGAAGAAACATTTTGA 540
 Db 500 TGCAGCTACTGCTGTGTGTCAAAAATTCACACAAAATCAAACTTGAAGAAACATTTTGA 559
 QY 541 AGCAAAACATGAAATTAACAAAACATATATATGAGTTTGAAGTGTGAAGAGC 600
 Db 560 AGCAAAACATGAAATTAACAAAACATATATATGAGTTTGAAGTGTGAAGAGC 619
 QY 601 CTTTAAAGAACTCAGCGCTGAAATTCATCATGTCAGCATATCAATGAAACCTCTATT 660
 Db 620 CTTTAAAGAACTCAGCGCTGAAATTCATCATGTCAGCATATCAATGAAACCTCTATT 679
 QY 661 CAAGTGTACCCAGGAGATGTGGAAACATTTGATCACCAGACAGCTGAAAGACA 720
 Db 680 CAAGTGTACCCAGGAGATGTGGAAACATTTGATCACCAGACAGCTGAAAGACA 739
 QY 721 TGGCAAGGCGCCAGGAGTGTGTGTCAAAAAGATTTCTTTGTGTGGCAAAAACATG 780
 Db 740 TGGCAAGGCGCCAGGAGTGTGTGTCAAAAAGATTTCTTTGTGTGGCAAAAACATG 799
 QY 781 GAGGAACTTCTGAACATGTGAGAGAAACCATAAAGAGAAATACTATGTGAAGTATG 840
 Db 800 GAGGAACTTCTGAACATGTGAGAGAAACCATAAAGAGAAATACTATGTGAAGTATG 859
 QY 841 CCGGAAACATTTAAAGCAAAAGATTAACCTTAAGCAACATGAAAGCAATGCCCCAGA 900
 Db 860 CCGGAAACATTTAAAGCAAAAGATTAACCTTAAGCAACATGAAAGCAATGCCCCAGA 919
 QY 901 AAGGATGTATGTGCTGTCCAGAGAGAGTGTGAAGAACCTATACTACTGTGTAA 960
 Db 920 AAGGATGTATGTGCTGTCCAGAGAGAGTGTGAAGAACCTATACTACTGTGTAA 979
 QY 961 TGTCCAAAGCCATATCTCTCTCTCCATGAGGAAAGCGCCCTTTGTGTGGAACATGC 1020
 Db 980 TGTCCAAAGCCATATCTCTCTCTCTCCATGAGGAAAGCGCCCTTTGTGTGGAACATGC 1039
 QY 1021 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAAGGCAATGCTGTGTGATGA 1080
 Db 1040 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAAGGCAATGCTGTGTGATGA 1099
 QY 1081 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAAGGCAATGCTGTGTGATGA 1140
 Db 1100 TGGCTGTGGCAAAACATTTGCAATGAAACAAAGTCTCAAGGCAATGCTGTGTGATGA 1159
 QY 1141 CTCCTATCTCACTGTATATCTCTCCCAAAAGGAAACAAAGGCAAGCTTATCTTTGTG 1200
 Db 1160 CTCCTATCTCACTGTATATCTCTCCCAAAAGGAAACAAAGGCAAGCTTATCTTTGTG 1219

QY 1201 TCMAAAGGAGAGTACCCAACTGTGTGGAAGACAAAGATGCTCTGACAGTTGCACTACT 1260
 Db 1220 TCMAAAGGAGAGTACCCAACTGTGTGGAAGACAAAGATGCTCTGACAGTTGCACTACT 1279
 QY 1261 TACCCCTGGCTAA 1273
 Db 1280 TACCCCTGGCTAA 1292

RESULT 14
 US-09-724-676A-44601
 : Sequence 44601, Application US/09724676A
 : GENERAL INFORMATION:
 : APPLICANT: Comugen LTD
 : TITLE OF INVENTION: Variants of alternative splicing
 : FILE REFERENCE: 129181.4 Comugen
 : CURRENT FILING DATE: 2000-11-28
 : NUMBER OF SEQ ID NOS: 97222
 : SOFTWARE: PatentIn version 3.2
 : SEQ ID NO 44601
 : LENGTH: 2384
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 US-09-724-676A-44601

Query Match 99.6%; Score 1268.2; DB 5; Length 2384;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCGAGAGAGGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
 Db 20 ATGCGAGAGAGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 79
 QY 61 GTGCGGCGCGCGCGGAGGTTTACAGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
 Db 80 GTGCGGCGCGCGCGGAGGTTTACAGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 139
 QY 121 AGGTGTCGTGGACGTGGCAGCGCGCTGGCCCTGGGCTTGGAGCGCGCGCGCTTCCGGAAGT 180
 Db 140 AGGTGTCGTGGACGTGGCAGCGCGCTGGCCCTGGGCTTGGAGCGCGCGCGCTTCCGGAAGT 199
 QY 181 TCCGCGCGCGCGGTGTGGCGGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 240
 Db 200 TCCGCGCGCGCGGTGTGGCGGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 259
 QY 241 ACCCGCGGAGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 300
 Db 260 ACCCGCGGAGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 319
 QY 301 CTCCTTCCCTGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 360
 Db 320 CTCCTTCCCTGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 379
 QY 361 CAAGCAGCGCGGAGACATTTTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 420
 Db 380 CAAGCAGCGCGGAGACATTTTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 439
 QY 421 CAGGAGTACATGTGGAGCGCGCACATTTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 480
 Db 440 CAGGAGTACATGTGGAGCGCGCACATTTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 499
 QY 481 TGCAGCTACTGCTGTGTGTCAAAAATTCACACAAAATCAAACTTGAAGAAACATTTTGA 540
 Db 500 TGCAGCTACTGCTGTGTGTCAAAAATTCACACAAAATCAAACTTGAAGAAACATTTTGA 559
 QY 541 AGCAAAACATGAAATTAACAAAACATATATATGAGTTTGAAGTGTGAAGAGC 600
 Db 560 AGCAAAACATGAAATTAACAAAACATATATATGAGTTTGAAGTGTGAAGAGC 619
 QY 601 CTTTAAAGAACTCAGCGCTGAAATTCATCATGTCAGCATATCAATGAAACCTCTATT 660

Db	620	CTTTAAAGAACATCAGCAGCTGATAAATCCATCGATGCGCAGATACGCATGAACTCATTT	679
Qy	661	CAAGTGTACCCAGGAAGATGTGTGGAAACACTTTTGATTCATCCACGCAAGCTGAAACGACA	720
Db	680	CAAGTGTACCCAGGAAGATGTGTGGAAACACTTTTGATTCATCCACGCAAGCTGAAACGACA	719
Qy	721	TGCGCAAGCCGACGAGGCGCTATGTATGTCAAAAAGAGATTTCTTTGTGCAAAAACATG	780
Db	740	TGCGCAAGCCGACGAGGCGCTATGTATGTCAAAAAGAGATTTCTTTGTGCAAAAACATG	729
Qy	781	GAGCGAAGCTTCTGAACATCTGTAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG	840
Db	800	GAGCGAAGCTTCTGAACATCTGTAGAGAAACCCATTAAGAGAAATACTATGTGAAGTATG	859
Qy	841	CCGGAANAACATTTAAACGCAAGATTTACCTTAAGCAGACATGAAAACTCATGCCCCAGA	900
Db	860	CCGGAANAACATTTAAACGCAAGATTTACCTTAAGCAGACATGAAAACTCATGCCCCAGA	919
Qy	901	AAGGATGTATGTGCGCTGTCCAAAGAGAAAGGCGTGGAGAAGCTATACCTATGTTGTTAA	960
Db	920	AAGGATGTATGTGCGCTGTCCAAAGAGAAAGGCGTGGAGAAGCTATATCACTATGTTTAA	979
Qy	961	TCTTCAAAGCCATATCTCTCTCTCCATGTAGGAAAGCCGCTTTTGTGTGTGAATGC	1020
Db	980	TCTTCAAAGCCATATCTCTCTCTCCATGTAGGAAAGCCGCTTTTGTGTGTGAATGC	1039
Qy	1021	TGCGTGTGGCAAAACATTTTGCAATGAAGAAAGTCTCACTAGGCAATGCTGTGTACATGA	1080
Db	1040	TGCGTGTGGCAAAACATTTTGCAATGAAGAAAGTCTCACTAGGCAATGCTGTGTACATGA	1099
Qy	1081	TCTGTACAGAGAAAGAAATGAGCTCAAAAGTCAAAAATCTCTGTGAANAACGAGATTGGC	1140
Db	1100	TCTGTACAGAGAAAGAAATGAGCTCAAAAGTCAAAAATCTCTGTGAANAACGAGATTGGC	1159
Qy	1141	CTCTCATCTCAGTGTATATCCCTCCCAAAGAGAAAGGCAAGGCTTATCTTTGTG	1200
Db	1160	CTCTCATCTCAGTGTATATCCCTCCCAAAGAGAAAGGCAAGGCTTATCTTTGTG	1219
Qy	1201	TCAAAAGGAGATCAACCCCACTGTGTGTGAAGACAAAGTCTCTGCACAGTTGCAGTACT	1260
Db	1220	TCAAAAGGAGATCAACCCCACTGTGTGTGAAGACAAAGTCTCTGCACAGTTGCAGTACT	1279
Qy	1261	TACCCCTTGCTTAA 1273	
Db	1280	TACCCCTTGCTTAA 1292	
RESULT 15			
US-09-724-676-44667			
: Sequence 44667, Application US/09724676			
: GENERAL INFORMATION:			
: APPLICANT: Compugen LTD			
: TITLE OF INVENTION: Variants of alternative splicing			
: FILE REFERENCE: 129181.4 Compugen			
: CURRENT APPLICATION NUMBER: US/09/724.676			
: CURRENT FILING DATE: 2000-11-28			
: NUMBER OF SEQ ID NOS: 97222			
: SOFTWARE: PatentIn version 3.2			
: SEQ ID NO 44667			
: LENGTH: 2536			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
US-09-724-676-44667			
Query Match 99.6%; Score 1268.2; DB 5; Length 2536;			
Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1270; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
Qy	1	ATGCGCAGCAGCGGCGCCGACGCGGCGCGCTCTCTGTGTGACCGGCGCGCTCCCGGAAGT	60
Db	20	ATGCGCAGCAGCGGCGCCGACGCGGCGCGCTCTCTGTGTGACCGGCGCGCTCCCGGAAGT	79
Qy	61	GTGCGGCGGTGCGCGCAGAGTTTACGACAGGAAGCCGTGCGGCGCGCGCGGTTCGCCGC	120

D	80	GTCCGCGCCGCGCGGAAGTTTCAGCAGGAGACCTGGGCGCGGCGCGCGGTATCCCGGC	139
Q	121	ACGTGTCTCGACACTGGACACGCGCGCTGGCCCTGGGCTTGGAGGCGCGGCGCCCTGGA	180
D	140	ACGTGTCTCGACACTGGACACGCGCGCTGGCCCTGGGCTTGGAGGCGCGGCGCCCTGGA	199
Q	181	TCGCGCGGCGCGTGGTGGCGCGGTGGGTGTCGTCTGACATCTGCGCGACGCTTCACTTGC	240
D	200	TCGCGCGGCGCGTGGTGGCGCGGTGGGTGTCGTCTGACATCTGCGCGACGCTTCACTTGC	259
Q	241	AGCGGCGAGAGCTCAGCTGCACCCGCGCGCGCGCGCGCTTCCAGAGGTTCACTG	300
D	260	AGCGGCGAGAGCTCAGCTGCACCCGCGCGCGCGCGCGCTTCCAGAGGTTCACTG	319
Q	301	CTCTCTTCCCTGACTCAGCGCGCAATTACAGCAAGCTTGGAAAGCTTACGCGACCTG	360
D	320	CTCTCTTCCCTGACTCAGCGCGCAATTACAGCAAGCTTGGAAAGCTTACGCGACCTG	379
Q	361	CAAGCACACGCGGCGAGAGACCATTGTGTTGTGACTATAGAGGTTGTGGCAAGCCTTCA	420
D	380	CAAGCACACGCGGCGAGAGACCATTGTGTTGTGACTATAGAGGTTGTGGCAAGCCTTCA	439
Q	421	CAGGACTACCTCTGAGCGCCGACATTCTGACTACACAGAGAAAGCCGTTTGTG	480
D	440	CAGGACTACCTCTGAGCGCCGACATTCTGACTACACAGAGAAAGCCGTTTGTG	499
Q	481	TGCAGCCACTGGCTGTGATCAAAATTCAACACAAATCAACTTTGAAGAAACATTTTGA	540
D	500	TGCAGCCAAATGGCTGTGATCAAAATTCAACACAAATCAACTTTGAAGAAACATTTTGA	559
Q	541	ACGCAAAACATGAAATCAACAAAGAAACATATATATGACGTTTTGACACTGTAGAGAC	600
D	560	ACGCAAAACATGAAATCAACAAAGAAACATATATATGACGTTTTGACACTGTAGAGAC	619
Q	601	CTTTAAGAAACATCAAGCGTGAATTCATGATGCCAGATACCAATTAACACTTAT	660
D	620	CTTTAAGAAACATCAAGCGTGAATTCATGATGCCAGATACCAATTAACACTTAT	679
Q	661	CAAGTATCCAGAGAAAGATGTGGGAAACATTTGCATCACCGACGAAGCTGAAACGACA	720
D	680	CAAGTATCCAGAGAAAGATGTGGGAAACATTTGCATCACCGACGAAGCTGAAACGACA	739
Q	721	TGCGAAGGCGCCAGGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACAG	780
D	740	TGCGAAGGCGCCAGGGCTATGTATGTCAAAAAGATGTTCTTTGTGGCAAAAACAG	799
Q	781	GAGGAACTTCTGAACACATGTGAGAGAAACCCATAAAGAGAAATACATATGTGAATAG	840
D	800	GAGGAACTTCTGAACACATGTGAGAGAAACCCATAAAGAGAAATACATATGTGAATAG	859
Q	841	CGGAAACATTTAAAGCAAAAGATTACCTTAAGACAACATGAAACCTATGCCGACAA	900
D	860	CGGAAACATTTAAAGCAAAAGATTACCTTAAGACAACATGAAACCTATGCCGACAA	919
Q	901	AAGGATGTATGTGCTGTCCAGAGAGAGGCTGTGAAGAACATATCTACTGTGTTAA	960
D	920	AAGGATGTATGTGCTGTCCAGAGAGAGGCTGTGAAGAACATATCTACTGTGTTAA	979
Q	961	TCGCAAAACCATATCTCTCTTCATGAGAAAGCGCCCTTTGTGTGAACATGC	1020
D	980	TCGCAAAACCATATCTCTCTTCATGAGAAAGCGCCCTTTGTGTGAACATGC	1039
Q	1021	TGGCTGTGGCAAAACATTTGCATGAACAAAGCTCACTATGAGCATCTGTGTATATGA	1080
D	1040	TGGCTGTGGCAAAACATTTGCATGAACAAAGCTCACTATGAGCATCTGTGTATATGA	1099
Q	1081	TCCTGACGAAGAAATGAAGCTCAAGCAAAAGAAATCTGTCAAAAACGAGGTTTGGC	1140
D	1100	TCCTGACGAAGAAATGAAGCTCAAGCAAAAGAAATCTGTGTCAAAAACGAGGTTTGGC	1159
Q	1141	CTCTCATCTCAAGTATATATCCCTCCCAAAAGAAACAGGCGAAGCTTATCTTGTG	1200

Db 1160 CTCTCATCTCACTGATATATCCCTCCCAAGGAAACAAGGCAAGGCTTATCTTTGTG 1219
QY 1201 TCAAAACGGAGAGTCACCCCACTGTGTGAAGACAAAGATGCTCTCGACAGTTGCAGTACT 1260
|||
Db 1220 TCAAAACGGAGAGTCACCCCACTGTGTGAAGACAAAGATGCTCTCGACAGTTGCAGTACT 1279
QY 1261 TACCCTTGAGCTAA 1273
|||
Db 1280 TACCCTTGAGCTAA 1292

Search completed: February 10, 2003, 13:06:19
Job time : 242.015 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:48:16 ; Search time 3381.18 Seconds
(without alignments)
10957.076 Million cell updates/sec

Title: US-09-831-426C-3

Sequence: 1 atgcgcagcagcgcgcgca.....cagctactacccttgctaa 1273

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database : GenEmbl1.*
1: gb_da.*
2: gb_htg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: gb_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_iny.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1233.2	96.9	1381	9 HUMTITA	D32257 Human GTF3A
2	1231.6	96.7	1399	6 E10962	E10962 cDNA's enco
3	1088.8	85.5	1173	9 HSU20272	U20272 Human DNA/R
4	919.2	72.2	1098	9 HSU14134	U14134 Human trans
5	889.8	69.9	1920	9 AK057993	AK057993 Homo sapi
6	719.2	56.5	1298	10 AF391799	AF391799 Mus muscu
7	691	54.3	1071	10 AF391798	AF391798 Rattus no
8	560.6	44.0	996	10 BC032292	BC032292 Mus muscu
9	470.2	36.9	97979	9 AC004739	AC004739 Homo sapi
10	470.2	36.9	193126	9 AC006045	AC006045 Homo sapi
11	452.6	35.6	111722	9 AC074390	AC074390 Homo sapi
12	452.6	35.6	206622	2 AC016444	AC016444 Homo sapi
13	447	35.1	199517	9 AL353741	AL353741 Human DNA
14	418.4	32.9	439	6 AX396111	AX396111 Sequence
15	416.8	32.7	439	6 AX396474	AX396474 Sequence
16	385.4	30.3	163421	2 AC012664	AC012664 Homo sapi
17	385.4	30.3	197709	9 AC073641	AC073641 Homo sapi
18	376	29.5	166141	9 AL137059	AL137059 Human DNA
19	373.4	29.3	200885	2 AC012460	AC012460 Homo sapi
20	372.4	29.3	1693	9 AF265440	AF265440 Homo sapi
21	361.4	28.4	1518	5 XELTFITIA	K02938 X. laevis 5S
22	355	27.9	1377	5 XBTF3A	X17695 Xenopus bor
23	350.8	27.6	1331	5 XELFINAB	M32472 X. borealis
24	340.6	26.8	1297	5 RPTREITIA	X58369 R. pipiens m
25	340.6	26.8	1305	5 RANFEITIA	M85211 Rana pipien
26	319.2	25.1	123905	9 AC116311	AC116311 Homo sapi
27	319.2	25.1	157165	2 AC104112	AC104112 Homo sapi
28	319.2	25.1	171007	2 AC021381	AC021381 Homo sapi
29	319.2	25.1	171705	2 AC113377	AC113377 Homo sapi
30	284.6	22.4	1314	5 BAFETIIA	AC021381 Homo sapi
31	278.6	21.9	67190	2 AC117566	X58366 B. americanu
32	234	18.4	234	6 AX396395	AC117566 Mus muscu
33	221.6	17.4	1129	5 IPOOF3A	AX396395 Sequence
34	202	15.9	162724	2 AC121381	268499 I. punctatus
35	202	15.9	182326	2 AC125693	AC121381 Rattus no
36	197.6	15.3	112902	9 AC005230	AC125693 Rattus no
37	177	13.9	272	6 AX322436	AC005230 Homo sapi
38	163	12.8	167342	2 AC124828	AX322436 Sequence
39	163	12.8	217225	10 AL513345	AC124828 Mus muscu
40	152.4	12.0	660	6 AX202580	AL513345 Mouse DNA
41	152.4	12.0	947	6 AX053174	AX202580 Sequence
42	152.4	12.0	947	6 AX059770	AX053174 Sequence
43	152.4	12.0	947	6 AX201765	AX059770 Sequence
44	152.4	12.0	947	6 AX201936	AX201765 Sequence
45	152.4	12.0	995	6 AX053173	AX201936 Sequence

ALIGNMENTS

RESULT 1
LOCUS HUMTITA
DEFINITION Human GTF3A mRNA for xenopus transcription factor IITA homologue,
complete cds.
ACCESSION D32257.1 GI:1000446
VERSION D32257.1
KEYWORDS GTF3A; Xenopus transcription factor IITA homologue.
SOURCE Homo sapiens cDNA to mRNA, clone lib:library of T.Fujiwara, S.Shin
and Y.Nakamura clone:39H11.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1381)
AUTHORS Arakawa,H., Nagase,H., Hayashi,N., Ogawa,M., Nagata,M.,

TITLE Fujiwara, T., Takahashi, E., Shin, S. and Nakamura, Y.
Molecular cloning, characterization, and chromosomal mapping of a
novel human gene (GTF3A) that is highly homologous to Xenopus
transcription factor IIA
JOURNAL Cytogenet. Cell Genet. 70 (3-4), 235-238 (1995)
MEDLINE 95309028
REFERENCE 2 (bases 1 to 1381)
AUTHORS Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (22-JUL-1994) Yusuke Nakamura, Cancer Institute,
Department of Biochemistry, 1-37-1 Kami-Ikebukuro, Toshima-ku,
Tokyo 170, Japan (E-mail: nakamura@genaxi.jifer.or.jp,
Tel: 03-3918-0111 (ex.4501), Fax: 03-3918-0342)
location/Qualifiers

FEATURES

source

1. 1381
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="39H11"
/clone_1lb="Library of T. Fujiwara, S. Shin and Y. Nakamura"
/gene="GTF3A"
1. 1381
/gene="GTF3A"
20. 1291
/gene="GTF3A"
/codon_start=1
/product="Xenopus transcription factor IIA homologue"
/protein_id="BA06988.1"
/db_xref="GI:1616942"

gene

CDS

BASE COUNT

386 a 350 c 354 g 291 t

ORIGIN

Query Match 96.9%; Score 1233.2; DB 9; Length 1381;
Best Local Similarity 99.5%; Pred. No. 4.1e-259;

Matches 1268; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

1 ATGGCGACACGGGCGGCGAGCGGCGGCTGCTGTGACCGCGCGGCTTCCCGAAGT 60
|||||
20 ATGGCGACACGGGCGGCGAGCGGCGGCTGCTGTGACCGCGCGGCTTCCCGAAGT 79
|||||
61 GTGCGGCGGTGCGCGAAGGTTTACAGAGGAGCGGTGGCGCGCGCGGTTCCCGC 120
|||||
80 GTGCGGCGGTGCGCGAAGGTTTACAGAGGAGCGGTGGCGCGCGCGGTTCCCGC 138
|||||
121 ACGTGTTCGGGACGCTGGGACGCGGCTGCGGCTTGGGCTTGAAGCGCGCGCTTGA 180
|||||
139 ACGTGTTCGGGACGCTGGGACGCGGCTGCGGCTTGGGCTTGAAGCGCGCGCTTGA 198
|||||
181 TCCGCGCGCGCTGGTTCGCGAGTTCGCTGCTTGAACATTCGCGAGCGCTTCATTGC 240
|||||
199 TCCGCGCGCGCTGGTTCGCGAGTTCGCTGCTTGAACATTCGCGAGCGCTTCATTGC 258
|||||
241 AGCGCGGAGAGCTCAGCTCCGAGCCCGCGCGCGCGGCTTCCCGAAGGTTTATGTG 300
|||||
259 ACCCGCGAGAGCTCAGCTCCGAGCCCGCGCGCGCGGCTTCCCGAAGGTTTATGTG 318
|||||
301 CTCCTTCCCTGACTGACGCGGCAATTCACAGCAAAAGCTTGAAGCTTGAAGCGCACCTTGTG 360
|||||
319 CTCCTTCCCTGACTGACGCGGCAATTCACAGCAAAAGCTTGAAGCTTGAAGCGCACCTTGTG 378
|||||
361 CAGGACACGGGGGAGAGACATTTGTTTGTGACTATGAAGGTTGTGGCAAGGCTTATAT 420
|||||
379 CAGGACACGGGGGAGAGACATTTGTTTGTGACTATGAAGGTTGTGGCAAGGCTTATAT 438
|||||
421 CAGGACATACATCTGAGCGCGCAATTCGACTCAGACAGAGAGAAAGCCGTTTGTGTTG 480
|||||
439 CAGGACATACATCTGAGCGCGCAATTCGACTCAGACAGAGAGAAAGCCGTTTGTGTTG 498
|||||

QY 481 TCCAGCACTGGCTGTGTGATCAAAAATTCAACACAAAATCAAACTTGAAGAAACATTTTGA 540
|||||
DB 499 TCCAGCACTGGCTGTGTGATCAAAAATTCAACACAAAATCAAACTTGAAGAAACATTTTGA 558
|||||
QY 541 ACCGAAACATGAAATTAACAAACAAATATATATGACAGTTTGAAGACTGTAAACAAAC 600
|||||
DB 559 ACCGAAACATGAAATTAACAAACAAATATATATGACAGTTTGAAGACTGTAAACAAAC 618
|||||
QY 601 CTTTAAAGAACATCAGACGCTGAAATTCATCAGTCCGACGATACCAATGAACCTGTAT 660
|||||
DB 619 CTTTAAAGAACATCAGACGCTGAAATTCATCAGTCCGACGATACCAATGAACCTGTAT 678
|||||
QY 661 CAAAGTACCCAGAGAGATGTGGGAACACTTGTGATCACCAGCAAGCTGAAGAGCA 720
|||||
DB 679 CAAAGTACCCAGAGAGATGTGGGAACACTTGTGATCACCAGCAAGCTGAAGAGCA 738
|||||
QY 721 TCCCAAGGCCCAAGGCTATGTATGTCAAAAAGATGTTCTTTTGGCAAAAACATG 780
|||||
DB 739 TCCCAAGGCCCAAGGCTATGTATGTCAAAAAGATGTTCTTTTGGCAAAAACATG 798
|||||
QY 781 GACGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 840
|||||
DB 799 GACGAACTTCTGAACATGTGAGAGAAACCCATTAAGAGAAATCTATGTGAAGTATG 858
|||||
QY 841 CCGGAAACATTTTAAAGCAAACTTACCTTAAAGCAACATGAATACTCATGCCCGCA 900
|||||
DB 859 CCGGAAACATTTTAAAGCAAACTTACCTTAAAGCAACATGAATACTCATGCCCGCA 918
|||||
QY 901 AAGGATGTATGTGCTGTCCAGAGAGAGCTGTGAAGAACCTTACTACTGTGTAA 960
|||||
DB 919 AAGGATGTATGTGCTGTCCAGAGAGAGCTGTGAAGAACCTTACTACTGTGTAA 978
|||||
QY 961 TCTCCAAAGCCATATCTCTCTCTTCCATGAGGAAAGCCGCTTTTGTGTGAACATGC 1020
|||||
DB 979 TCTCCAAAGCCATATCTCTCTCTTCCATGAGGAAAGCCGCTTTTGTGTGAACATGC 1038
|||||
QY 1021 TGGCTGTGGCAAAACATTTTGCATGAACAAAGCTTCACTAGGATGCTGTGTGATGA 1080
|||||
DB 1039 TGGCTGTGGCAAAACATTTTGCATGAACAAAGCTTCACTAGGATGCTGTGTGATGA 1098
|||||
QY 1081 TCCCTGACAAAGAAATATGAAGCTCAAGATCAAAAATCTCGTGAAGAAAC-AGATTTGG 1139
|||||
DB 1099 TCCCTGACAAAGAAATATGAAGCTCAAGATCAAAAATCTCGTGAAGAAAC-AGATTTGG 1158
|||||
QY 1140 CCTCTCATCTCAGTGAATATATCTCTCCAAAAGAAAGAGGAGGCTTATCTTGT 1199
|||||
DB 1159 CCTCTCATCTCAGTGAATATATCTCTCCAAAAGAAAGAGGAGGCTTATCTTGT 1217
|||||
QY 1200 GTCAAAAGGAGAGTCAACCAAGCTGTGGAAGAAAGATGCTCTGACAGTGTGATAC 1259
|||||
DB 1218 GTCAAAAGGAGAGTCAACCAAGCTGTGGAAGAAAGATGCTCTGACAGTGTGATAC 1277
|||||
QY 1260 TTACCTTGGCTAA 1273
|||||
DB 1278 TTACCTTGGCTAA 1291
|||||

RESULT 2
E10962
LOCUS E10962 1399 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA's encoding hTFIIIA protein.
ACCESSION E10962
VERSION E10962.1 GI:22028822
KEYWORDS JP 1996070870-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1399)
AUTHORS Fujiwara, T., Takeda, S., Shimada, Y., Ozaki, K. and Shin, T.
TITLE HUMAN TRANSCRIPTIONAL FACTOR IIA GENE
JOURNAL Patent: JP 1996070870-A 1 19-MAR-1996;

COMMENT	OTSUKA PHARMACEUT CO LTD
OS	Homo sapiens (human)
PN	JP 1996070870-A/1
PD	19-MAR-1996
PF	05-SEP-1994 JP 1994211022
PI	FUJIMURA TSUTOMU, TAKEDA SEI, SHIMADA YOSHIKAZU, OZAKI KOICHI,
PI	SHIN TEIKIN
PC	C12N15/09//A61K31/70,A61K48/00,C07H21/04;
CC	strandedness: Single;
CC	topology: Linear;
Key	Location/Qualifiers
FM	
FM	
PT	Source
PT	1..1399
PTM	/organism='Homo sapiens'
PT	1..139
PT	5'UTR
PT	20..1288
PT	CDS
PT	/product='hNF11A protein'
PT	1289..1399.
PT	3'UTR
Source	Location/Qualifiers
	1..1399
	/organism='Homo sapiens'
	/db_xref='taxon:9606'
BASE COUNT	405 a 349 c 354 g 291 t
ORIGIN	

	Query Match:	96.7%;	Score 1231.6;	Df: 6;	Length 1399;	
	Best Local Similarity:	99.5%;	Pred. No. 9.2e-259;			
	Matches 1267;	Conservative	0;	Mismatches 4;	Indels 3;	Gaps 3.
OY	1	ATGCGCAGACAGCGCGCCGCACGCGCGCGCGTCCCTGCTGGTGACCGGCCGCCCGCTCCCGCAACT	60			
Db	20	ATGCGCAGCAGCGCGCGCCGCACGCGCGCGCGTCCCTGTGTGACCGCCGCCCGCTCCCGCAACT	79			
OY	61	GTCGCGGCGCTGCGCGCGAAGTTTCAGCAGAGGACCCGTGGGCGCGGCGCGCGGTTCGCGCG	120			
Db	80	GTCGCGGCGCTGCGCGCGAAGTTTCAGCAGAGGACCCGTGGGCGCGGCGCGCGGTTCGCGCG	138			
OY	121	ACGTGTCTCCGCGCACGTGCGCAGCGCGCTTGCGCTTGAGGCGCGCGCCCGCCCTGGA	180			
Db	139	ACGTGTCTCGGCGACGTGCGCAGCGCGCTTGCGCTTGAGGCGCGCGCCCGCCCGCCCTGGA	198			
OY	181	TCCCGCGCGCGTGTGCGCGAGTCCGTGCTGCTTGACATCCGCACAGCGGTTCATTGC	240			
Db	199	TCCCGCGCGCGTGTGCGCGAGTCCGTGCTGCTTGACATCCGCACAGCGGTTCATTGC	258			
OY	241	AGCGCGGGAAGCGTCACGTCTCCGACCCCCCGCGCGCGCGCTTCCCAGAGAATTCACTG	300			
Db	259	AGCGCGGGAAGCGTCACGTCTCCGACCCCCCGCGCGCGCGCTTCCCAGAGAATTCACTG	318			
OY	301	CTCCTTCCCTGACTGACGCGCCAAATTACAGCAAACCTGGAAGCTTAGCGCGCACCTGTG	360			
Db	319	CTCCTTCCCTGACTGACGCGCCAAATTACAGCAAACCTGGAAGCTTAGCGCGCACCTGTG	378			
OY	361	CACACACAGCGGAGAGAACATTSTTTGTGACTATGAAGGGGTGGCAAGGCGCTTCAT	420			
Db	379	CACACACACGCGGAGAGAACATTSTTTGTGACTATGAAGGGGTGGCAAGGCGCTTCAT	438			
OY	421	CAGGACTACCATCTGAGCGGCCACATTCTGACTCACACAGAGAAAAGCCGTTTGTTC	480			
Db	439	CAGGACTACCATCTGAGCGGCCACATTCTGACTCACACAGAGAAAAGCCGTTTGTTC	498			
OY	481	TGCAGCCACTGGCGTGTGATCAAAAANTTCACACAATAATCAAACTTGAAGAAACATTTTCGA	540			
Db	499	TGCAGCCCAATGGCGTGTGATCAAAAANTTCACACAATAATCAAACTTGAAGAAACATTTTCGA	558			
OY	541	ACGGAACATGAAAAATCAACAAAAACAATATATATGACGTTTTGAAGACGTGAAGAAGAC	600			
Db	559	ACGGAACATGAAAAATCAACAAAAACAATATATATGACGTTTTGAAGACGTGAAGAAGAC	618			
OY	601	CTTTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCAGCATCCCAATGAACCTCATTT	660			
Db	619	CTTTAAGAAACATCAGCAGCTGAAAAATCCATCAGTGCAGCATCCCAATGAACCTCATTT	678			

QY	661	CAATGTATACCCAGGAAGATGTGGGAAACCTTTGTCATCCACGACAGCTGAACGACA	720
Db	679	CAATGTATACCCAGGAAGATGTGGGAAACCTTTGTCATCCACGACAGCTGAACGACA	738
QY	721	TGCCAAGCCGACGAGGCGTATGTATGTCAAAAAGAGATGTCTTTGTGGCAAAAACATG	780
Db	739	TGCCAAGCCGACGAGGCGTATGTATGTCAAAAAGAGATGTCTTTGTGGCAAAAACATG	798
QY	781	GAGGACCTCTCAACATGTGAGGAAACCCATTAAGAGAAATCTATGTCAACTATG	840
Db	799	GAGGACCTCTCAACATGTGAGGAAACCCATTAAGAGAAATCTATGTCAACTATG	858
QY	841	CCGGAACAACATTTAAACGCAAGATTACCTTAAGCAACACATGAAAACATCATGCCCCAGA	900
Db	859	CCGGAACAACATTTAAACGCAAGATTACCTTAAGCAACACATGAAAACATCATGCCCCAGA	918
QY	901	AAGGATGTATGTGCGTGTCCAAAGAGCGTGTGGAAGCAATATCTCTGTGTTAA	960
Db	919	AAGGATGTATGTGCGTGTCCAAAGAGCGTGTGGAAGCAATATCTCTGTGTTAA	978
QY	961	TCTCCAAAGCCATATCTCTCTCTCCATGAGGAAAGCCGCTTTTGTGTGAACATGC	1020
Db	979	TCTCCAAAGCCATATCTCTCTCTCCATGAGGAAAGCCGCTTTTGTGTGAACATGC	1038
QY	1021	TGGCGTGGCAAAAACATTTGCAATGCAAGTCTCACTAGGCATCTTTGTACATGA	1080
Db	1039	TGGCGTGGCAAAAACATTTGCAATGCAAGTCTCACTAGGCATCTTTGTACATGA	1098
QY	1081	TCTGCACAAAGAAAATGAAGTCAAAAGTCAAAAATCTGTGTAAAAC -GGAGTTGG	1139
Db	1099	TCTGCACAAAGAAAATGAAGTCAAAAGTCAAAAATCTGTGTAAAACGGAGTTGG	1158
QY	1140	CCCTCATCTCACTAGTGGATATATCCCTCCCAAAAAGAAACAAGGCAAGCTTATCTTTGT	1199
Db	1159	CCCTCATCTCACTAGTGGATATAT -CCCTCCCAAAAAGAAACAAGGCAAGCTTATCTTTGT	1217
QY	1200	GTCAAAACGGAGAGTACCCCACTGTGTGGAAGACAAGATCTCTGCAAGTTGACATTC	1259
Db	1218	GTCAAAACGGAGAGTACCCCACTGTGTGGAAGACAAGATCTCTCTGCAAGTTGCAATTC	1277
QY	1260	TTACCCCTTGGCGTAA 1273	
Db	1278	TTACCCCTTGGCGTAA 1291	

```

RESULT 3
HSU20272
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

HSU20272
Human DNA/RNA-binding protein mRNA, partial cds.
U20272
U20272.1 GI:644870

1173 bp
mRNA
linear
PRI 26-SEP-1995

Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (sites)
Brew,P.D., Nagle,J.W., Canning,R.D., Ozato,K., Biddison,W.E. and
Drew,K.G.
Cloning and expression analysis of a human cDNA homologous to
Xenopus FFIIIA
Gene 159 (2), 215-218 (1995)
7622052
2 (bases 1 to 1173)
Becker,K.G.
Direct Submission
Submitted (25-JAN-1995) Kevin G. Becker, Nat. Inst. of Neur.
Diseases and Stroke/NIH, Neuroimmunology Branch, 9000 Rockville
Pike, Bethesda, MD 20892, USA
location/Qualifiers
1..1173
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="C2H2-34.10"
/sex="female"
/tissue_type="brain; hippocampus"
/clone_lib="Stratagene #936205"
/dev_stage="2 years old"
<1..1092
/note="similar to Xenopus TfIIia encoded by Swissprot
Accession Number P03001"
/product="DNA/RNA-binding protein"
/protein_id="AA05623.1"
/db_xref="GI:644871"
/translation="PPAVVAESVSLTIADAFIAGSSAPTPRPALPRFICSPD
CSANYSKAWKLDALHLCIKTGERPFYCDYEGCGKAFIDYHLISHLIHTHGEKPFVCA
NGCDOKFNKSNLKKHFERKHENOKOYICSEFDCKTFKKHLOLKHOCQHTNEPLF
KCTGEGCGHAPSPKLRHAKAHGEGYCKGKSPVAKTKTELLKHVHEKKEILCE
VCKTQPKRRDYLKOHMKTHAPRDRGRCRRCGCRRTYTPENQSHLSHESRPFV
CEHAGCGKTPAKKOSLTRHAYVHDPDKKKMLKVKREKRSILASHLSGITIPKRNQ
QGLSCQKESPEVCVEDKMLSTVAVLTIG"

misc_feature
118..184
/note="encodes zinc-finger"
misc_feature
208..274
/note="encodes zinc-finger"
misc_feature
298..367
/note="encodes zinc-finger"
misc_feature
394..460
/note="encodes zinc-finger"
misc_feature
484..550
/note="encodes zinc-finger"
misc_feature
565..631
/note="encodes zinc-finger"
misc_feature
649..709
/note="encodes zinc-finger"
misc_feature
736..805
/note="encodes zinc-finger"
misc_feature
829..895
/note="encodes zinc-finger"
BASE COUNT 363 a 283 c 265 g 262 t
ORIGIN

Query Match 85.5% Score 1088.8 DB 9: Length 1173:
Best Local Similarity 99.8% Pred No. 1.5e-227;
Matches 1090: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 182 CCGCGCGCGTGGTCCGAGTGGTGTCTCTTGACCATGCGCGACCGCTTATTCGA 241
DB 1 CCGTCTGCTGTGTGCGAGTGGTGTCTCTGACCATGCGCGACCGCTTATTCGA 60
OY 242 GCGGCGAGAGCTAGCTCCGACCGCGCGCGCGCGCTTCCGAGAGTTTCATGCG 301
DB 61 GCGGCGAGAGCTAGCTCCGACCGCGCGCGCGCGCGCTTCCGAGAGTTTCATGCG 120
OY 302 TCTTCTGCTAGCTCAAGCGCAATTACAGCAAAAGCTGAGCTTACGCGCACCTGTGC 361
DB 121 TCTTCTGCTAGCTCAAGCGCAATTACAGCAAAAGCTGAGCTTACGCGCACCTGTGC 180
OY 362 AAGCAGCGGGGAGAGACCAATTGTTGTGACTATGAGGGGTGGCAAGGCGTTTCATC 421
DB 181 AAGCAGCGGGGAGAGACCAATTGTTGTGACTATGAGGGGTGGCAAGGCGTTTCATC 240
OY 422 AGGAGCTACCATCTGAGCGCGCACATTTGACTGCACAGAGGAGAAAGCGTTTGTGT 481
DB 241 AGGAGCTACCATCTGAGCGCGCACATTTGACTGCACAGAGGAGAAAGCGTTTGTGT 300
OY 482 GCAGCGACTGCTGTGATCAAAATTCAACAAATTCAAACCTGAGAAACATTTTGA 541
DB 301 GCAGCGACTGCTGTGATCAAAATTCAACAAATTCAAACCTGAGAAACATTTTGA 360
OY 542 CGCAAAATGAAATCAACAAACATATATATGAGTTTGAAGACTGTAAAGAGCC 601
DB 361 CGCAAAATGAAATCAACAAACATATATATGAGTTTGAAGACTGTAAAGAGCC 420

OY 602 TTTAAGAAATCAGACGCTGAAATTCATGATGCGCAGCATACCAATGAACCTCTATTG 661
DB 421 TTTAAGAAATCAGACGCTGAAATTCATGATGCGCAGCATACCAATGAACCTCTATTG 480
OY 662 AAGTGTACCCAGAGAGATGTGGAAACACTTTGCATGACCCAGCAAGCTGAAAGCAT 721
DB 481 AAGTGTACCCAGAGAGATGTGGAAACACTTTGCATGACCCAGCAAGCTGAAAGCAT 540
OY 722 GCCAAGGCCAGCAGGGGCTATGATGTCAAAAAGATGTCTTTGGGCAAAAACATGG 781
DB 541 GCCAAGGCCAGCAGGGGCTATGATGTCAAAAAGATGTCTTTGGGCAAAAACATGG 600
OY 782 ACGGAACCTTGAAACATGTGAGAGAAACCCATTAAGAGGAATACTATGTGAAGTATG 841
DB 601 ACGGAACCTTGAAACATGTGAGAGAAACCCATTAAGAGGAATACTATGTGAAGTATG 660
OY 842 CGGAAACATTTAAAGCCAAAGATTAACCTTAAGCAACACATGAAGAAATCATGCCGAG 901
DB 661 CGGAAACATTTAAAGCCAAAGATTAACCTTAAGCAACACATGAAGAAATCATGCCGAG 720
OY 902 AGGATGTATGTGCTGTGCCAGAGAGGCTGTGAGAGACCTATACTATGTGTATAT 961
DB 721 AGGATGTATGTGCTGTGCCAGAGAGGCTGTGAGAGACCTATACTATGTGTATAT 780
OY 962 CTCGAAAGCCATATCTCTGCTTCATGAGAGAGGCGCGCTTTGTGTGAGACATGCT 1021
DB 781 CTCGAAAGCCATATCTCTGCTTCATGAGAGAGGCGCGCTTTGTGTGAGACATGCT 840
OY 1022 GCGTGTGCAAAACATTTGCATGAATGAACAAAGCTTCATGACATGCTGTTGTATGAT 1081
DB 841 GCGTGTGCAAAACATTTGCATGAATGAACAAAGCTTCATGACATGCTGTTGTATGAT 900
OY 1082 CCGTGAAGAAAGAAATGAAGCTCAAAAGTCAAAAATTCGTAAGAAAGAGATTGGCC 1141
DB 901 CCGTGAAGAAAGAAATGAAGCTCAAAAGTCAAAAATTCGTAAGAAAGAGATTGGCC 960
OY 1142 TCTCATCTCAGTGGATATATCTCTCCCAAGGAAACAAGGCAAGCTTATCTTGTGT 1201
DB 961 TCTCATCTCAGTGGATATATCTCTCCCAAGGAAACAAGGCAAGCTTATCTTGTGT 1020
OY 1202 CAAAAGGAGAGTCAACCACTGTGTGAAGACAAAGATGCTGACAGTTCAGTACTT 1261
DB 1021 CAAAAGGAGAGTCAACCACTGTGTGAAGACAAAGATGCTGACAGTTCAGTACTT 1080
OY 1262 ACCCTGGCTAA 1273
DB 1081 ACCCTGGCTAA 1092
RESULT 4
HSD14134 1098 bp mRNA linear PRI 01-OCT-1994
LOCUS HSD14134
DEFINITION Human transcription factor TfIIA (TFIIIA) mRNA, partial cds.
ACCESSION U14134
VERSION U14134.1 GI:551534
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1098)
Drew, P.D., Nagle, W.W., Camling, R.D., Ozato, K., Biddison, W.E. and
Becker, K.G.
TITLE Cloning and expression analysis of a cDNA encoding human TfIIIA
JOURNAL Unpublished
AUTHORS Becker, K.G.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-1994) Kevin G. Becker, National Institute of
Neurological Diseases and Stroke/NIH, Neuroimmunology Branch, 9000
Rockville Pike, Bethesda, MD 20892, USA
FEATURES
source location/Qualifiers 1..1098

[illegible]

Db	1408	TAAACGAAAGATTACCTTAAGCAACACATGATAAAAGTCATGCCCAAGAAAGGATGTATG	1467
Qy	913	TCGCTGTCCAAAGAAAGGCTGTGGAAAGAACTATATCTACTGTGTTTAATCTCCAAAGCA	972
Db	1468	TCGCTGTCCAAAGAAAGGCTGTGGAAAGAACTATATCAACTGTGTATTAATCTCCAAAGCA	1527
Qy	973	TATCTCTCTCTTCATGAGGAAGAGCGGCCCTTTTGTGTGTGACATGCTGGCTGTGGCA	1032
Db	1528	TATCTCTCTCTTCATGAGGAAGAGCGGCCCTTTTGTGTGTGACATGCTGGCTGTGGCA	1587
Qy	1033	AACATTTGCATGAAACAAAGCTCTACTAGGATATGCTGTGTATGATCTCTGACAAGAA	1092
Db	1588	AACATTTGCATGAAACAAAGCTCTACTAGGATATGCTGTGTATGATCTCTGACAAGAA	1647
Qy	1093	GAATATGAAGCTCAAGTCAAAAAATCTCTGGAAGAAACGAGATTTGGCCTTCATCTCAG	1152
Db	1648	GAATATGAAGCTCAAGTCAAAAAATCTCTGGAAGAAACGAGATTTGGCCTTCATCTCAG	1707
Qy	1153	TGGATATATCCCTCCCAAGGAACAAAGGCAAGCGTTATCTTTGTGTCAAAACGGAGA	1212
Db	1708	TGGATATATCCCTCCCAAGGAACAAAGGCAAGCGTTATCTTTGTGTCAAAACGGAGA	1767
Qy	1213	GTCACCCAACTGTGTGGAAAGACAGATGCTCTGCAGAGTTGCAGTACTTACCTTGGCTA	1272
Db	1768	GTCACCCAACTGTGTGGAAAGACAGATGCTCTGCAGAGTTGCAGTACTTACCTTGGCTA	1827
Qy	1273	A 1273	
Db	1828	A 1828	
RESULT 6			
LOCUS	AF391799	1298 bp	MRNA linear ROD 31-JAN-2002
DEFINITION	Mus musculus transcription factor I1IA mRNA, partial cds.		
ACCESSION	AF391799		
VERSION	AF391799.1	GI:18448381	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
REFERENCE	1 (bases 1 to 1298)		
AUTHORS	Hans,J.S., Hocker,J.R., Cheng,Y.G., Lerner,M.R., Brackett,D.J., Lightfoot,S.A., Hanes,R.J., Madhusudan,K.T. and Moreland,R.J.,		
TITLE	CDNA cloning, DNA binding, and evolution of mammalian transcription factor I1IA		
JOURNAL	Gene 282 (1-2), 43-52 (2002)		
MEDLINE	21673987		
PUBMED	11814676		
REFERENCE	2 (bases 1 to 1298)		
AUTHORS	Hans,J.S., Hocker,J.R., Lerner,M.R., Brackett,D.J., Lightfoot,S.A., Hanes,R.J., Madhusudan,K. and Moreland,R.J.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-JUN-2001) Biochemistry/Molecular Biology, University of Oklahoma Health Science Center, 800 Research Parkway, Suite 448, Oklahoma City, OK 73104, USA		
FEATURES			
source			
	1..1298		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	<1..1203		
	/note="F11IA: Cys2His2 zinc finger protein; similar to xenopus laevis transcription factor I1IA Swiss-Prot Accession Number P03001"		
	/codon_start=1		
	/product="transcription factor I1IA"		
	/protein_id="AA69686.1"		
	/db_xref="GI:18448382"		
	/translation="EEETGVOSI.RAPVACPVSRGRHVAAPROCPDPRVALPEPVSA EAVSIRIADAPVAGACGAPRPAALPSRFTSPDCASINAKAKIDAHICKITGR PFOVDIECGKARITRDHLSRHLITGKPEVCAADCCNOKFWTKSLUKHIERKRG NPOVDICESTEGCKAKRKHQDILRTHQCSHTSEPLERCTHDECGKHPASPSRLRKHK		


```
repeat_region      8718. .9750  
/rpl_family="L1"  
repeat_region      9751. .12094  
/rpl_family="L1"  
repeat_region      12095. .12372  
/rpl_family="Alu"  
repeat_region      12373. .12436  
/rpl_family="(CA)n"  
repeat_region      12437. .12842  
/rpl_family="L1"  
repeat_region      12888. .13001  
/rpl_family="L1"  
repeat_region      13003. .13358  
/rpl_family="MERL"  
repeat_region      13365. .14214  
/rpl_family="L1"  
repeat_region      14264. .14417  
/rpl_family="L1"  
repeat_region      14415. .14729  
/rpl_family="L1"  
repeat_region      14816. .15026  
/rpl_family="MERL_type"  
repeat_region      15648. .15729  
/rpl_family="(CATA)n"  
repeat_region      16041. .16362  
/rpl_family="L1"  
repeat_region      16679. .16732  
/rpl_family="L1"  
repeat_region      17025. .17315  
/rpl_family="Alu"  
repeat_region      17465. .17685  
/rpl_family="L1"  
repeat_region      18044. .18633  
/rpl_family="L1"  
repeat_region      18630. .18750  
/rpl_family="L1"  
repeat_region      18760. .19497  
/rpl_family="L1"  
repeat_region      19508. .19692  
/rpl_family="L1"  
repeat_region      19702. .19814  
/rpl_family="L1"  
repeat_region      21491. .21543  
/rpl_family="L2"  
repeat_region      21987. .24424  
/rpl_family="L1"  
repeat_region      24425. .24720  
/rpl_family="Alu"  
repeat_region      24722. .24821  
/rpl_family="L1"  
repeat_region      25383. .25745  
/rpl_family="L1"  
repeat_region      29202. .29272  
/rpl_family="MERL_type"  
repeat_region      30338. .30386  
/rpl_family="(CA)n"  
repeat_region      30984. .31045  
/rpl_family="MERL_type"  
repeat_region      31041. .31398  
/rpl_family="MERL_type"  
repeat_region      32178. .32319  
/rpl_family="MIR"  
repeat_region      32475. .32515  
/rpl_family="L2"  
repeat_region      32589. .32646  
/rpl_family="L2"  
repeat_region      32940. .33224  
/rpl_family="Alu"  
misc_feature  
complement(33558. .33668)  
repeat_region      33843. .33929  
/note="match to EST AA928509 (MID:g3076800) om17907.s1"  
/rpl_family="MERL_type"  
repeat_region      35891. .36381
```

```
repeat_region      37910. .38123  
/rpl_family="MERL_type"  
repeat_region      38302. .38768  
/rpl_family="L1"  
repeat_region      42711. .43026  
/rpl_family="L1"  
repeat_region      43023. .43418  
/rpl_family="L1"  
repeat_region      43552. .43806  
/rpl_family="L1"  
repeat_region      43847. .44539  
/rpl_family="L1"  
repeat_region      44534. .45153  
/rpl_family="L1"  
repeat_region      45132. .45245  
/rpl_family="L1"  
repeat_region      45246. .45667  
/rpl_family="MERL"  
repeat_region      45674. .45869  
/rpl_family="L1"  
repeat_region      45871. .46089  
/rpl_family="Alu"  
repeat_region      46091. .47111  
/rpl_family="L1"  
repeat_region      47186. .47354  
/rpl_family="MERL_type"  
repeat_region      47529. .47577  
/rpl_family="MERL_type"  
repeat_region      47673. .48037  
/rpl_family="L2"  
repeat_region      48067. .48494  
/rpl_family="L2"  
repeat_region      51739. .51788  
/rpl_family="(CA)n"  
repeat_region      52362. .52665  
/rpl_family="L1"
```

Query Match 36.9%; Score 470.2; DB 9; Length 97979;
Best Local Similarity 76.8%; Pred. No. 4,9e+92;
Matches 655; Conservative 0; Mismatches 158; Indels 40; Gaps 5;

```
QY 369 TGTGACTATGAAGGGTGTGCGAAGCGCTTCATCAGGAGTACCATCTGACGCCACATT 448  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79437 TGTGACTATGAAGGGTGTGCGAAGCGCTTCATCAGGAGTACCATCTGACGCCACATC 79378  
QY 449 CTGACTCAGCAGGAGAAAGCGTTGTTGTGCGACCGACGCGCTGTATCAAA----- 503  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79377 CTGATTCACATGTGAGAAAGCGTTTATTTGTGCGACGTAGCGCTGTATCAAAAAAAA 79318  
QY 504 -----AATTCACACAAATCAAACTTGAGAAACATTTTACGCCAAACATGAAATC 557  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
DB 79317 AAAAAAATCCCAAGAAATCCAACTGGAAGAAATTTGCATGCAAAACATGAAAAATC 79258  
QY 558 AACAAAAACATATATATGACGTTTGAAGACTGTAGAGACCTTTAAGAAATCAGC 617  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79257 TGCAAAACCATATGATGTGCAATTTTGAAGAGATGAGAAACCTTTAAGAAAGCATCGC 79198  
QY 618 AGCTGAAATCATCATGCGCAGCATACCAATGAACTTATTTCAAGTGTACCAAGAG 677  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79197 AGCTGAAATCATCATGCGCAGCATGAAATTAACCCCTATTTCAAGTTTAACTTGAAAG 79138  
QY 678 GATGTGGAACACTTGTGACCCAGCAAGCTGAAGCAAGATGCCAAAGCCCAAGAG 737  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79137 GATGTGGAACACTTGTGTTTCAACCTCCAGCGTGGAGACAGATGCAAGTTTCATCAGA 79078  
QY 738 GCTATGTATGTCAAAAAGATGTTCCGTTTGTGGCAAAAACATGACGGAACCTTGGAAC 797  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79077 AATTTATATGTCAATGAGATGTTCTTTGTGGCAAAAACATGAGC----- 79032  
QY 798 ATGTGAGAGAACCCATTAAGAGAAATACTATGTGAAGTATGCGGGAACATTTAAAC 857  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
DB 79031 ---ACAGAGAAATCCATTAAGAAAGAAATA--ACATGATGTATGCCAAGAAACATTTCAAT 78977
```

QY	858	GCAAGATTACCTTAAGCAACATGAAACACATGACCCCGAAGAGGATGTATGTCGT	917
Db	78976	GTAAGATTTATCTCAATATACCTGTGAAAACACTAGCCTCGAAGAGATGTATGTCAC	78917
QY	918	GTCGAGAGAGAGCGCTGTGACAGACCTTACTACTGTGTTTAATCTCCAAAGCCATATCC	977
Db	78916	ATCCAGAGAGAGAGCTCCACAGAGA-----ACGTAAATTAATCTCCAAAGCGTCTTC	78866
QY	978	TCTCTTCATGTAGGAGAACCGCCCTTTTGTGTGTGAACATGCTGCGCTGTGGCAAAACAT	1037
Db	78865	TTCCTTTCTTAAGGAGAGAGACACCATTATATGTGAAACATGCTGCGCTGTGGCAAAAGT	78806
QY	1038	TTGCAATGAAGAAACAATCATCTAGCAAGCGTGTGTATCATGATCCGTGAACAACAAAAA	1097
Db	78805	TGCGTATGAAGAACAACTCTCACAAGACATGCTATTGTGTATGAGCTTGACAGAGAAAA	78746
QY	1098	TGAAGTCGAAGTCAAAAAATCTCTGAAAAACGGAGTTGGCCCTCATCTCAGTGCAT	1157
Db	78745	TAAAGATCAACAAGTAAGCTATCTCTGAAAAATCAAAATTTGGCCCTCATCTCAGT-AAA	78687
QY	1158	ATATCCCTCCCAAAAGAGAAACAGGCAAGCGTTATCTTTGTGTCAAAAGCAGAGATCAC	1217
Db	78686	ATATCCCTCGTAAGAGAGAAACAAGAACGAGTTTATCTTTGCATTTGAATGAAGGAGATCAT	78627
QY	1218	CCAAGCTGTGCGA 1230	
Db	78626	TGAATTGTGTATTGA 78614	

RESULT 10					
LOCUS	AC006045/c				
DEFINITION	Homo sapiens clone	NH0547615,	complete sequence.		
ACCESSION	AC006045				
VERSION	AC006045.2	GI:4753227			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 193126)				
JOURNAL	Waterston,R.H.				
REFERENCE	The sequence of Homo sapiens clone				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 193126)				
JOURNAL	Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (22-NOV-1998) Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
JOURNAL	MO 63108, USA				
REFERENCE	3 (bases 1 to 193126)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-MAY-1999) Genome Sequencing Center, Washington				
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,				
AUTHORS	MO 63108, USA				
TITLE	On May 5, 1999 this sequence version replaced gi:3907455.				
JOURNAL	Location/Qualifiers				

```

SOURCE
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0547615"
BASE COUNT 63568 a 34097 c 33661 g 61800 t
ORIGIN

```

Query Match:	36.9%	Score 470.2:	DB 9,	Length 193126:
Best Local Similarity:	76.8%	Pred. No. 5,4e-97:		
Matches 655:	Conservative	0:	Mismatches 158:	Indels 40:
			Gaps	5:
0Y	389	TGTACATATAAGGGGTGTGGCAGAGCCTTCATCAGGACACTACATCTGAGCCGCCACATT	448	
Db 128874		TGTACATATAAGGGGTGTGGCAGAGCCTTCATCAGGACACTACATCTGAGCCGCCACATT	128815	

OY	449	CTGCTCACACGAGAAAGGCCGTTGTTGTGACGACGCTGGCTGTGATCA-----	503
Db	128814	CTGATTCACATTGCGAAGAAAGCCTTTATTTGTGAGCTAGTGGCTGTGTATAAAGAAAA	128755
OY	504	-----AATTCAACACAAAAATCAACTGGAAGAAACATTTTGAACGCAAAACATGAAAATC	557
Db	128754	AAAAAGATATCCACAGAAATATCCATGGAAGAAACATTTTGGCATGCAAAACATGAATATC	128695
OY	558	AACAAAAACAATATATATGATGAGTTTGAAGACTGTAAAGAAAGCCTTTAAAGAAACATCATC	617
Db	128694	TGCAAAAGCAATATGATGATCAATTTTGAAGATGTAAAGAAAGACCTTTTAAGAGCATCAGC	128635
OY	618	AGCTGAAATATCATCATGTCGACAGCATACCAATGAAACCTTATTCAACTATCCAGCAAG	677
Db	128634	AGCTGAAATATCCATCATGTCGACAGCATGAAAPAAACCCCTTATCAAGTTAACTTTGAAG	128575
OY	678	GATGTGGGAACACTTTGATATACCCAGACAAAGCGAATGCAATGCCAGCCACAGGAC	737
Db	128574	GATGTGGAAGAACCTTTGTTTCAACCTCCAGGCTGGAGAGACAGATGACAGTTCACTAGA	128515
OY	738	GCTATGTATGTCAAAAGAGATGTCCTTTGTGTGGCAAAACATGAGACGGAACCTTCTGAAC	797
Db	128514	AAATATATGTGCATATGGGATGTCTTTTGTGGGAGAAACATGATGAC-----	128469
OY	798	ATGTGATAGAGAAACCCATTAAGAGGAAATACATATGGAATGTGGCCGGAAGAACTTTAAAC	857
Db	128468	---AGAGAGAAATCTCATAGAAAGGAAAT--ACATGATGTTATGCCAGAAACATTCAAAAT	128414
OY	858	GCAAAATTTACCTTAAGCACACATGAAAACTCATGCCCCGAAAGGAGTATATGTGCGT	917
Db	128413	GTCGAAAGTATATCTCAATATACCTGTGAAAACTCATGTCTCGAAGAGATGTATGTCAAC	128354
OY	918	GTCGAAAGAGGCTGTGGAAGAACCTTATACGTACGTGTTTAAATCTCAAGACATATCC	977
Db	128353	ATCCAGAGAGAGCTGCAGANAGA-----ACGTATTTAAATCTCCAGAGCTGTGCTTC	128303
OY	978	TCTCTTCATGAGGAGAAAGCCGCCCTTTTGTGTGTGAACATGATGTGGCTGTGGCAAAACAT	1037
Db	128302	TTTCTTTCTTAAAGGAGAAAGCACACCATTTATATGTGTGAACATGCTGGGCTGTGGCAAAAGGT	128243
OY	1038	TTTCAATGAAGAACAAATCTCAGTACGCAATGCTGTGTATCATGATCTCTGCAACAGAGAAA	1097
Db	128242	TGCGTATGAAGAACAACTCTCCAGACATGCTATTTGTCTATGACCTTGACAGAGAGAAAA	128183
OY	1098	TGAAGCTCAAGTCAAAAAATCTCGTGAAGAAACGAGTTTGGCCTCTCATCTCAGTGGAT	1157
Db	128182	TAAAGATCAACAGTAAAGCTATCTCTGAAAAATGAAATTTGGCCTCTCATCTCAGT-AAA	128124
OY	1158	ATATCCCTCCCAAAAGAGAAACAGGCAAGGCTTATCTTTTGTGCAAAACGAGAGCTCAC	1217
Db	128123	ATATCCCTCGTAAAGAGAAACAGAACGAGTTTATCTTTTTCATGTAAATGGAGATCAT	128064
OY	1218	CCAAGTGTGTGGA 1230	
Db	128063	TGAATGTGTATTTGA 128051	

RESULT	11
AC074390/c	
LOCUS	AC074390 111722 bp DNA linear PRI 21-FEB-2002
DEFINITION	Homo sapiens BAC clone RP11-S15K14 from 7, complete sequence.
ACCESSION	AC074390
VERSION	AC074390.6 GI:18250153
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Sulston,J.E. and Waterston,R. Toward a complete human genome sequence Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE	
TITLE	
JOURNAL	

MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 111722)
 AUTHORS Bielicki,L., Haakenson,W. and Boyer,E.
 TITLE The sequence of Homo sapiens BAC clone RP11-515K14
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 111722)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 111722)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (19-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 111722)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Jan 19, 2002 this sequence version replaced gi:17976493.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0515K14

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR/>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is CTD-2105K18, 2000 bp overlap; the clone sequenced to the right is RP11-533K11, 2000 bp overlap. Actual start of this clone is at base position 141753 of CTD-2105K18; actual end is at base position 16829 of RP11-533K11. Polymorphisms have been identified between AC016444 and AC074390.

FEATURES Data from AC016444 was used to finish this clone, AC074390.
 source Location/Qualifiers

1..111722
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"
 /clone="RP11-515K14"
 /clone_1lb="RPCI-11"
 1..249
 /rpt_family="MaLR"
 repeat_region
 /rpt_family="MaLR"
 250..640
 /rpt_family="MaLR"
 repeat_region
 1650..2014
 /rpt_family="ERV1"
 repeat_region
 2015..3233
 /rpt_family="L1"
 repeat_region
 3234..3578
 /rpt_family="ERV1"
 repeat_region
 3586..3835
 /rpt_family="ERV1"
 repeat_region
 4560..4873
 /rpt_family="ERV1"
 repeat_region
 4880..4975
 /rpt_family="GA-rich"
 repeat_region
 4991..5133
 /rpt_family="ERV1"
 repeat_region
 5577..5801
 /rpt_family="L1"
 repeat_region
 6687..6982
 /rpt_family="Alu"
 repeat_region
 7392..7449
 /rpt_family="AT-rich"
 repeat_region
 7495..7564
 /rpt_family="AT-rich"
 repeat_region
 7559..8728
 /rpt_family="L1"
 repeat_region
 8746..9038
 /rpt_family="Alu"
 repeat_region
 9017..9053
 /rpt_family="(TAA)n"
 repeat_region
 9061..9362
 /rpt_family="Alu"
 repeat_region
 9372..10143
 /rpt_family="L1"
 repeat_region
 9934..9963
 /rpt_family="(TC)n"
 repeat_region
 10144..10454
 /rpt_family="Alu"
 repeat_region
 10427..10454
 /rpt_family="(A)n"
 repeat_region
 10455..11140
 /rpt_family="L1"
 repeat_region
 11487..11960
 /rpt_family="L2"
 repeat_region
 12112..12305
 /rpt_family="L2"
 repeat_region
 12306..12615
 /rpt_family="Alu"
 repeat_region
 12591..12615
 /rpt_family="(A)n"
 repeat_region
 12616..12766
 /rpt_family="L2"
 repeat_region
 13024..13088
 /rpt_family="(TA)n"
 repeat_region
 13613..13641
 /rpt_family="AT-rich"
 repeat_region
 14281..14328
 /rpt_family="(TA)n"
 repeat_region
 14329..15028
 /rpt_family="MER2_type"
 repeat_region
 15050..15494
 /rpt_family="MER2_type"

```

repeat_region      15445. .15464
                    /rpt_family="T)n"
repeat_region      15495. .16698
                    /rpt_family="L1"
repeat_region      17544. .17927
                    /rpt_family="L1"
repeat_region      19584. .19618
                    /rpt_family="MaLR"
repeat_region      21646. .21769
                    /rpt_family="MIR"
repeat_region      21860. .22043
                    /rpt_family="MER1_type"
repeat_region      22044. .22427
                    /rpl_family="MaLR"
repeat_region      22428. .22467
                    /rpt_family="MER1_type"
repeat_region      23300. .23664
                    /rpt_family="MaLR"
repeat_region      23870. .23910
                    /rpt_family="(tRNA)n"
repeat_region      24405. .24635
                    /rpl_family="L1"
repeat_region      24636. .24901
                    /rpl_family="MaLR"
repeat_region      25481. .25504
                    /rpt_family="AT_rich"
repeat_region      26543. .26672
                    /rpt_family="MER1_type"
repeat_region      26769. .26911
                    /rpl_family="CR1"
repeat_region      26928. .27295
                    /rpt_family="MaLR"
repeat_region      27653. .27678
                    /rpt_family="AT_rich"
repeat_region      27934. .28093
                    /rpt_family="MER1_type"
repeat_region      29167. .29194
                    /rpl_family="AT_rich"
repeat_region      29191. .29309
                    /rpt_family="Alu"
repeat_region      29348. .29644

```

Query Match	35.68;	Score 452.6;	DB 9;	Length 111722;
Best Local Similarity	80.38;	Pred. No. 3.5e-8;		
Matches 685;	Conservative	0;	Mismatches 139;	Indels 29;
			Gaps	12;

QY	269	CCGCGCCCGCGGCTCCCGAGAGGTTCAATGCTCTCCCTCGATCGACCCCAATTTC	328
Db	87601	CCCCCTACACAACTTCTTAGAGGGGTACTGCTACTTCCC-CAGTGTGCACTTAC	87543
QY	329	AGCAAAAGCTGGAGCTTGACCGCAGCTGTGCAACAGACAGGGGAGAGACCATTTGTT	388
Db	87542	AAAAAGGCTGGAGAGTTTAAATGTCAGCCTGTG--AATCACTGGGGAAGAAACCATTTGTT	87485
QY	389	TGTGACTATGAAGGTTGTGGCAGGCGCTTCATC-AGGAGCAACCATCTGCAGCCGACAT	447
Db	87484	TATGACTGTGAAGGGTGTGGCAAGACCCAGCTCAAGAAGAGCTAACATCTGAGTTGTCACTG	87425
QY	448	TTGTGCTCACACAGGAGAAACCCGTTGTTGTCGAGCCACTGCTGTGATC-AAAAAT	506
Db	87424	CCCTGA-TCATACCTGGCAGACACCCCATATCTTTTGACAGCTAGTGGCTGATCTTAAAAAT	87366
QY	507	TCAACACAAAATCAAACTTGAAGAAACATTTTGAAGCCAAACATGTAAATCAACAAAAC	566
Db	87365	TCAACACAAAATCAAACTTGAAGAAACATTTTGAAGCCAAACATGTAAATCAACAAAAC	87306
QY	567	A-----ATATATTTGAGTTTGAAGACTGTAAAGAGCCTTTAAAGAAATCATGACGT	621
Db	87305	AAAAATATCATATGTCAATTTTGGAGGTTTGAAGAGAGCCTTTACG-AACATATGAGCGCT	87247
QY	622	GAAAATTCATCATGCTGGCAGCATATGCCAATGAACCTCTATTCAAGTTATCCCGAGGAAGATG	681
Db	87246	GAAAATTCATCATGCTGGCAGCATATGCCAATGAACCTCTATTCAAGTTATCCCGAGGAAGATG	87187

OY	582	TGGGAAACATTGTGCATCACCGACGAAGCTGAAAGACATGCAAGGCGCCAGAGGGCTA	741
Db	87186	TGGGAACATTTTTCCTCCCATCCATGCACG-----AAGCGGCAAGGCCCATGCGGGCTC	87134
OY	742	TGTATGTCAAAAAGATGTTCTTGTGTGGCAAAACATGAGCGAACTTGTGAACATGT	801
Db	87133	TATATGTCAAAAAGCGCTGTTTCTTTGTGTGGTGAAGATGAAACAGATTTTGAACATGT	87074
OY	802	GAGGAAGACCATTAAGAGGAATTCCTATGTGAAGTATGCGGAAACATTTAACGCCAA	861
Db	87073	GAGGAAGACTTTAAANAAAGAAACACATATGAAGCAGCGCCAGAAACATTTAAAC----	87018
OY	862	AGATTACTCTTAAGCAACACATGAAACTCATGCCCCAGAAAGGATTTATGTGCTGTCC	921
Db	87017	--ATTATCTTAACCAACATNTGACACCTCATATGCCCCAAAGGATAT-TTTTGACATAC	86961
OY	922	AAGGAAGGCTGTGGAAGCAACCTATACTACTGTGTTTATCTCCAAAGCATATCCTCTC	981
Db	86960	AAGGAAGTCTCTGGAAGAACCTACCAACACTGTGTTTACTTACAAATTCATATCTCTTTC	86901
OY	982	CTTCATGAGGAAGCGCGCCCTTTGTGTGAACATCTGGCGTGGCAAAACATTTCG	1041
Db	86900	CAT--GAGAAAAAAGGTGCCATTTTGTGATGTGAACATTTTGGCTGTGGCAAAACATTTCG	86843
OY	1042	AATGAACAACAAGCTCTCACTAGCAGCATGCTGTTTACATCATGCTGCACAAGAAAAATGAA	1101
Db	86842	AATGAACAACAAGCTTCTTAGCGATCGTGTGTACAT-ATTGTGACAAGAAGAAATGAA	86784
OY	1102	GCTCAAGTCAAA 1114	
Db	86783	GCTCAGAGTATA 86771	

RESULT	12
LOCUS	AC016444
DEFINITION	AC016444 Homo sapiens chromosome 11 clone RP11-384I14 map 11. WORKING DRAFT
ACCESSION	AC016444
VERSION	AC016444.3 GI:10045722
KEYWORDS	HTGS; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 206622) Birren B., Linton L., Nisbaum C. and Lande E. Homo sapiens chromosome 11, clone RP11-384I14 Unpublished
AUTHORS	2 (bases 1 to 206622) Birren B., Linton L., Nisbaum C., Lande E., Allen N., Anderson M., Baldwin J., Barna N., Beckerly R., Boguslavsky L., Boukhgalter B., Brown A., Castle A., Colangelo M., Collins S., Collimore A., Cook P., Deatellano K., Dewar K., Domingo M., Donelan L., Doyle M., Ferrair P., Fitzhugh W., Forrest C., Funke R., Gage D., Gallagan J., Gardyna S., Grant G., Hages B., Heatford A., Horton L., Howland J. C., Johnson R., Jones C., Kann L., Karats A., Klein J., Lohoczky J., Lieu C., Locke K., Macdonald P., Marguis N., McEwan P., McGuirk J., McKernan K., McLaughlin J., Meldrim J., Meterson K., Pollara V., Riley R., Roy A., Santos R., Severy P., Stange-Thomann N., Stojanovic N., Subramanian A., Talmes J., Tesiaye-S., Tirrell A., Vassiliev H., Vo A., Wheeler J., Wu X., Wyman D., Ye W. J., Zimmer A. and Zody M.
TITLE	Direct Submission
JOURNAL	Submitted (26-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Sep 9, 2000 this sequence version replaced gl:6778507.
COMMENT	All repeats were identified using RepeatMasker: Smith A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html ----- Genome Center -----

Db	Accession	Version	Keywords	Source	Organism	Reference	Title	Journal	Comment
36836	GAGGAAAGCTATTAAAAAGCAACATATGAAAGCAGCCACGAAAACTTTAAAC								
36892	AGATTACCTTAAGCAACACATATAAAGCTATGCGCTGTC								
36949	AAAGGAAGTCTGGAAGAACTACACAAGTGTTCACATAATCATATCCTTTTC								
982	CTTTCATGAGGAAGCCGCCCTTTTGTGTGTGGAACATGCTGCTGTGGCAAAACATTTCG								
37009	CAT-GAGAAAAGAGGTGCGCATTTTGGCATGTGAACATGTGTGGCGTGGCAAAACATTTCG								
1042	AATGAAACAAGTCTACTAGCGCATGCTGTGTGATCATGATCTCTGACAGAGAAATGAA								
37067	AATGAAACAAGTCTTTTAGCGCATGCTGTGTGATCAT-ATTCTGACAGAGAAATGAA								
1102	GCTCAAGCTCAA	1114							
37126	GCTCAGAGTATA	37138							
RESULT 13	AL353741	199517 bp	DNA	linear	RP1-575C20				
LOCUS	AL353741								
DEFINITION	Human DNA sequence from clone RP1-575C20 on chromosome 9, complete								
ACCESSION	AL353741								
VERSION	AL353741.16	GI:12584694							
KEYWORDS	HTG.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.								
AUTHORS	Laird, G.								
JOURNAL	Direct Submission								
COMMENT	Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk requests: clonerequests@sanger.ac.uk On Jan 28, 2001 this sequence version replaced gi:12580980. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9 RP1-575C20 is from the library RPc1-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACe3.6 This sequence is the entire insert of clone RP1-575C20 The true left end of clone RP1-279B1 is at 116738 in this sequence. The true right end of clone RP1-21817 is at 84470 in this sequence. Location/Qualifiers 1..199517 /organism="Homo sapiens"								

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone IMAGE:5467230
 /clone_lib="NIH-MGC_41"
 /tissue_type="amelanotic melanoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pORF1; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACAGAG(c). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH-MGC Library.

Query Match	72.58;	Score 923;	DB 13;	Length 1078;
Best Local Similarity	95.28;	Pred. No. 3.8e-192;		
Matches 1026;	Conservative 0;	Mismatches 43;	Indels 9;	Gaps 7;

QY	66	GAGCTGGCCGAAAGGTTCTAAGACAGGAGACCCCTGGAGCCGAGCGCGGATTCGCCGACAGTG	1.25
Db	1	GCGCTGGCCGAAAGGTTAAGCAAGGAGACCGTGGGCCGAGCGCGCGGATTCGCCGACAGTG	60
QY	126	TCTGCGCAGTGCGGACAGCCGCTCTGGCTTGAGAGCGCCGCGCCCTGGATCCGC	18.55
Db	61	TCTGCGCAGTGCGGAGCCGCTCTGGCTTGAGAGCGCCGCGCCCTGGATCCGC	1.20
QY	186	CGGCGCTGGTCCCGCAGTGCAGTGCCTTTCAGCACTTCGCGCAGCGCGATTCATTCAGCGC	24.54
Db	121	CGGCGCTGGTCCCGCAGTGCAGTGCCTTTCAGCACTTCGCGCAGCGCGATTCATTCAGCGC	1.80
QY	246	GCGAGAGCTCACTCCGACCCCGCCGCGCGCGCTTCCAGAGAGTTCAATCTGCTCT	30.50
Db	181	GCGAGAGCTCACTCCGACCCCGCCGCGCGCGCTTCCAGAGAGTTCAATCTGCTCT	24.00
QY	306	TTCCTACTACGACGCCCAATTACAGCAAAACCTTGAAAGCTTACGCGCGCCCTGCGCAAGC	3.65
Db	241	TTCCTACTACGAGGCCCAATTACAGCAAAACCTTGAAAGCTTACGCGCGCCCTGTCGCAAGC	3.00
QY	366	ACACGGGGGAGAGACCAATTTCTTTGTGACATATGAGAGGGTGTGGCAAGCCCTTTCATCAGGG	4.25
Db	301	ACACGGGGGAGAGACCAATTTCTTTGTACACTATAGAGGGTGTGGCAAGCCCTTTCATCAGGG	3.60
QY	426	ACTACACATGTGAGCGCCGACATTTCTGTGACTACACAGAGAAAGCAAGCGCTTGTGTGAG	48.55
Db	361	ACTACACATGTGAGCGCCGACATTTCTGTGACTACACAGAGAAAGCGCTTGTGTGTGAG	4.20
QY	486	CCACTGGCTGTATCAAAAATTCAACACAAATCAAACTTGAAGAAACATTTTGAAGCA	54.55
Db	421	CCAAATGGCTGTATCAAAAATTCAACACAAATCAAACTTGAAGAAACATTTTGAAGCA	4.80
QY	546	AACATGAATCAACAAAAACATATATATATGACGTTTGTAGACATCTGAAGAGACCTTTA	60.50
Db	481	AACATGAATCAACAAAAACATATATATGACGTTTGTAGACATCTGAAGAGACCTTTA	54.00
QY	606	AGAATCATCAGCAGCTGAAATTCATCATGTCGACGATACCAATGAAACCTCTATTCAAGT	66.55
Db	541	AGAATCATCAGCAGCTGAAATTCATCATGTCGACGATACCAATGAAACCTCTATTCAAGT	60.00
QY	666	GTAACCCAGAGACGATGTGGAAACACTTTCATACACCGACGAAGCGTGAAGACGACATGCA	72.55
Db	601	GTAACCCAGAGAGATGTGGAAACACTTTCATACACCGACGAGAGCTGAAGACGACATGCA	66.00
QY	726	AGGCCACGAGGGGCTATGTATGTCAAAAAGAGTTCCTTTGTGGCAAAAACATGAGAGG	78.55
Db	661	AGGCCACGAGGGGCTATGTATGTCAAAAAGAGTTCCTTTGTGNCAAAACATGAGAGG	72.00
QY	786	AACCTTGAAACATGTGAGAGAAACCCATTAAGAGAAATACTATATGCAACTATATGCCGGA	84.55
Db	721	AACCTTGATGAACTGTGAGAGAAACCCATTAAGAGAAATACTATGTGAGATATGCCGGA	77.99

QY 845 AATCATTTAAACCAAGAATTTACCTTAAAGCAACACATGAAATCAAAATCATGCCCCCAAGAAAGG 905

Db 780 NAACATTTAAACGGCAAGATTTACCTTAAAGCAACACATGAAATCAAAATCATGCCCC -GAAAGG 838

QY 906 ATGTATGTGCTGTCCAGAGAGAGGCTGTGGAGAAAGCTATATACATGTTTATATCTCC 965

Db 839 ATGTATGTGCTGTCCAGAGAGAGGCTGTGGAGAAAGCTATATACAAATGTTTTATATCTCC 898

QY 966 AAAACCATTCATCCTCTCTCTTCATGAGAA -GCGCCCTTTGTGTGTGAACATGCT -G 1022

Db 899 -AAAGCATATCTCTCTCTTCATGAGAAAGAGCGCCCTTTTGGGTGTGAACATGCTGG 957

QY 1023 GCTGTGGCAAAACATTTGCAATGCAAAACAAAGTCTACAGCAATGCTGTTGTA -CATGA 1080

Db 958 CTGTGGGCAAAACATTTGGCATGTGAACCAAGTCTCCCTTAGGCATGCTGTTGTGAACATGAT 1017

QY 1081 TCTGTGACAGAGAGAAAT -GAAGCTCAAAAGTCAAAAATTTCTGTGTAATAAGAGATTT 1137

Db 1018 CCGTGCACAGAGAGAAATGGAATCTTCATATTCAAAAATTTTCGGGAAAAAAGCGGAT 1075

RESULT 2	AL518337	LOCUS	AL518337	958 bp	mRNA	linear	EST 13-FEB-2001
DEFINITION	AL518337	LM1_NFL011_NBCL	Homo sapiens cDNA clone CS0D0A091018 5				
ACCESSION	AL518337	prime, mRNA sequence.					
VERSION	AL518337.1	GI:12781830					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.						
JOURNAL	1 (bases 1 to 958)						
COMMENT	L.W.B., Gruber,C., Jéssée,J. and Polayes,D.						
	Full-length cDNA libraries and normalization						
	unpublished (2001)						
	Contact: Genoscope						
	Genoscope, Centre National de Séquençage						
	Bp 191 91006 EVRY cedex - France						
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.						

```

SOURCE
1. 958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GSDA009Y018"
/clone_1lb="LTI_NFL011_NBC1"
/sex="male"
/tissue.type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMWSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMWSPORT 6 vector. Library is not normalized, but is the control for the normalized libraries. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"
BASE COUNT
270 a 244 c 232 g 195 t 17 others
ORIGIN

```

	72.28%	Score 919;	DB 9;	Length 958;
Query Match	Similarity 97.5%;	Pred. No. 2, Be-191;		
Best Local Similarity	Conservative 16;	Mismatches 6;	Indels 2;	Gaps 2.
Matches 933;				
QY	76 GAGAGTTGACACCAGGAGACCGTGGGCCGGGGCCGTTCCTCCGGCACGTGTCGGGCACG	135		
Db	1 GAAGCTTGACACCAAGGAGCCGTGGGCCGGGGCCGGCGGTTCCCGACAGTGTCTCCTGGCAGCG	60		
OY	136 TGGCAGCGCCCGCTGGAGCCCTTGGAGAAGCCCGCGCCCTCTGGATCCGCGCGGCGCGTGA	195		


```
Db 601 AGCTGAACGACATGCCAAGCCGACGAGGGCTATGTATGTCAAAAAGGATTTCCCTTG 660
Qy 768 TGGCAAAAACATGACGCACTTGTGAACATGTGAGAGAAACCCATAAGAGAAATAC 827
Db 661 TGGCAAAAACATGACGCACTTGTGAACATGTGAGAGAAACCCATAAGAGAAATAC 720
Qy 828 TATGTAAGTATGCCGCAAAACATTTAAGCAAAAGATTACCTTAAGCAACATGAAAA 887
Db 721 TATGTAAGTATGCCGCAAAACATTTAAGCAAAAGATTACCTTAAGCAACATGAAAA 780
Qy 888 CTCATAGCCCCGAGAAAGGATGTATGCTGCTGCAAGAGAGAGCTGTGAAACCTTATA 947
Db 781 CTCATAGCCCCGAGAAAGGATGTATGCTGCTGCAAGAGAGAGCTGTGAAACCTTATA 840
Qy 948 CTACTGTGTTTATCTCCAAAGCCATATCCTCTCTCTTCATGAGAAACCCGCTTTTG 1007
Db 841 CAAGTGTGTTTATCTCCAAAGCCATATCCTCTCTCTTCATGAGAAACCCGCTTTTG 900
Qy 1008 TGTGTGAA 1015
Db 901 TGTGTGAA 908

RESULT 4
AL555777 881 bp mRNA linear EST 16-FEB-2001
LOCUS AL555777 LTI_NFL006_P12 Homo sapiens cDNA clone CS0DK002YJ18 5
DEFINITION prime, mRNA sequence.
ACCESSION AL555777
VERSION AL555777.1 GI:12897829
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 881)
AUTHORS L.L.W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YJ18"
/life_type="LTI_NFL006_P12"
/issue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
```

```
BASE COUNT 252 a 229 c 225 g 175 t
ORIGIN
Query Match 69.0%; Score 877.8; DB 9; Length 881;
Best Local Similarity 99.8%; Pred. No. 3e-182;
Matches 879; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 82 TCAGAGAGGAGCCGTGGCGCGCGCGGTTCCCGGACGCTGTCTGGCACCTGGCAG 141
Db 1 TCAGAGAGGAGCCGTGGCGCGCGCGGTTCCCGGACGCTGTCTGGCACCTGGCAG 60
Qy 142 CGCGCCTGGCCTGGCGGTTGGAGCGCGCGCGCTGATTCGCGCGCGCGCTGCTGCCGA 201
```

```
Db 61 CGCGCCTGGCCTGGCCTGGAGGCGCGCGCGCGCTGGATCCGCGCGCGCTGTCGCCGA 120
Qy 202 GTGCGTGTGCTGCTTTGACCATGCGCGAGCGCTTCATTGCAAGCGGAGAGCTCAGCTCC 261
Db 121 GTGCGTGTGCTGCTTTGACCATGCGCGAGCGCTTCATTGCAAGCGGAGAGCTCAGCTCC 180
Qy 262 GACCGCGCGCGCGCGCGCGCTTCCAGAGGTTCACTGCTGCTTCCCTGCACTGAGAGGC 321
Db 181 GACCGCGCGCGCGCGCGCGCTTCCAGAGGTTCACTGCTGCTTCCCTGCACTGAGAGGC 240
Qy 322 CAATTACAGCAAAAGCTTGAAGCTTGAAGCGCGCACCTGTGCAACACAGCGGGGAGAGAC 381
Db 241 CAATTACAGCAAAAGCTTGAAGCTTGAAGCGCGCACCTGTGCAACACAGCGGGGAGAGAC 300
Qy 382 ATTTGTTGTGACTGATGAGAGGTGAGCAAGGCTTCATCAGAGCACTACATCTGAGACG 441
Db 301 ATTTGTTGTGACTGATGAGAGGTGAGCAAGGCTTCATCAGAGCACTACATCTGAGACG 360
Qy 442 CCACATTCGACTCACACAGAGAAAGCGCTTTGTTGTGACGCCACTGCGTGATCA 501
Db 361 CCACATTCGACTCACACAGAGAAAGCGCTTTGTTGTGACGCCACTGCGTGATCA 420
Qy 502 AAAATTCAACACAAATCAACTGAGAAACATTTGAAAGCAACATGAAATCAACA 561
Db 421 AAAATTCAACACAAATCAACTGAGAAACATTTGAAAGCAACATGAAATCAACA 480
Qy 562 AAAACAATATATATGCACTTTTGAAGCTGTGAAGAGCTTTAAGAAACATCAGACCT 621
Db 481 AAAACAATATATATGCACTTTTGAAGCTGTGAAGAGCTTTAAGAAACATCAGACCT 540
Qy 622 GAAATTCATGAGTGCAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATG 681
Db 541 GAAATTCATGAGTGCAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Qy 682 TGGGAACACCTTGTGATCACCACAGCAAGCTGAAGCAATGCCAAGCCGACGAGGCTA 741
Db 601 TGGGAACACCTTGTGATCACCACAGCAAGCTGAAGCAATGCCAAGCCGACGAGGCTA 660
Qy 742 TGTATGTCAAAAAGGATGCTCTTGTGGCAAAAACATGAGAGCACTTGAACATGT 801
Db 661 TGTATGTCAAAAAGGATGCTCTTGTGGCAAAAACATGAGAGCACTTGAACATGT 720
Qy 802 GAGAGAAACCCATAAGAGCAATGATGTGAAGTATGCGGAAACATTTAAACGAA 861
Db 721 GAGAGAAACCCATAAGAGCAATGATGTGAAGTATGCGGAAACATTTAAACGAA 780
Qy 862 AGATTACCTTAAGCAACATGAAACATGATGATGATGATGATGATGATGATGATGATGATG 921
Db 781 AGATTACCTTAAGCAACATGAAACATGATGATGATGATGATGATGATGATGATGATGATG 840
Qy 922 AAGAGAGGCTGTGAGAGCACTTACTACTGCTTTAATC 962
Db 841 AAGAGAGGCTGTGAGAGCACTTACTACTGCTTTAATC 881

RESULT 5
BMS62682 1059 bp mRNA linear EST 20-FEB-2002
LOCUS BMS62682 5', mRNA sequence.
DEFINITION AGENCOURT_6588903 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478431
ACCESSION BMS62682
VERSION BMS62682.1 GI:13808978
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1059)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@pds-remail.nih.gov
```


Db	937	GTACTATGTAAGGCTGTGGCAAGCCCTTCATCAGGAGCTACCATCTGAGCGCACATTTC	878
Qy	450	TGACTCACACAGAGAAAAAGCCGTTTGTGTGTGCAGCCACTGGCTGTGATCAAAATTTCA	509
Db	877	TTATCTCACAGAGAAAAAGCCGTTTGTGTGTGC -GCCAATGCGCTGTGATCAAAATTTCA	819
Qy	510	ACACAAATTCAACTTGAAGAAACATTTTGAACGGCAACATGAAATTCACAAACAAAT	569
Db	818	ACACAAATTCAACTTGAAGAAACATTTTGAACGGCAACATGAAATTCACAAACAAAT	759
Qy	570	ATATATGCAATTTTGAAGAGCTGTAAAGAGACCTTTAAGAACAATCAGACAGCTAAATTC	629
Db	758	ATATATGCAATTTTGAAGAGCTGTAAAGAGACCTTTAAGAACAATCAGACAGCTAAATTC	699
Qy	630	ATCAGTCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGATGTGGGAAC	689
Db	698	ATCAGTCCAGCATACCAATGAACCTCTATTCAAGTGTACCCAGGAAGATGTGGGAAC	639
Qy	690	ACTTTCATACCCAGCAAGTGAAGAGCAATGCCAAGCCCAAGAGGCTATGTATGTC	749
Db	638	ACTTTCATACCCAGCAAGTGAAGAGCAATGCCAAGCCCAAGAGGCTATGTATGTC	579
Qy	750	AAAAAGATGTCCTTTGTGTGCAAAAACATGTGACGGAACCTTCTGAACATGTAGAGAAA	809
Db	578	AAAAAGATGTCCTTTGTGTGCAAAAACATGTGACGGAACCTTCTGAACATGTAGAGAAA	519
Qy	810	CCCTTAAGAGGAAATACTATGTGAAGTATGGCCGGAAGAACTTAAAGCGGAAGATTACC	869
Db	518	CCCTTAAGAGGAAATACTATGTGAAGTATGGCCGGAAGAACTTAAAGCGGAAGATTACC	459
Qy	870	TTTAAGCAACATGAAACATCATGCCGCCAGAAAGGATGTATGTGCGTGTCCAAGAGAG	929
Db	458	TTTAAGCAACATGAAACATCATGCCGCCAGAAAGGATGTATGTGCGTGTCCAAGAGAG	399
Qy	930	GCTGTGAAACAACTTACTACTGTGTTTAATCTCMAAGCAATATCTCTCCCTCCAG	989
Db	398	GCTGTGAAACAACTTACTACTGTGTTTAATCTCMAAGCAATATCTCTCCCTCCAG	339
Qy	990	AGCAAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGTGGCAAAACATTTGGCAATGAAC	1049
Db	338	AGCAAAAGCCGCTTTTGTGTGTGAACATGCTGGCTGTGGCAAAACATTTGGCAATGAAC	279
Qy	1050	AAAGTCTCAGTAGCAGTGCCTGTTGTACATGATCTCGACACAAAGAAATGAACCTCAAG	1109
Db	278	AAAGTCTCAGTAGCAGTGCCTGTTGTACATGATCTCGACACAAAGAAATGAACCTCAAG	219
Qy	1110	TCAAAAAATTCGTGTGAAGAAAGGAGATTTGGGCTCTCATCTCAGTGGAATATATCCCTCCA	1169
Db	218	TCAAAAAATTCGTGTGAAGAAAGGAGATTTGGGCTCTCATCTCAGTGGAATATATCCCTCCA	159
Qy	1170	AAAGGAACAGAGGAGGAGGCTATCTTTGTGTCAAAAGAGAGATCCACCACTGTGTGG	1229
Db	158	AAAGGAACAGAGGAGGAGGCTATCTTTGTGTGTCAAAAGAGAGATCCACCACTGTGTGG	99
Qy	1230	AACGACAAGATGCTCTGTGACAGTTTGCAGTACTTACCCCTTGGCTAA	1273
Db	98	AACGACAAGATGCTCTGTGACAGTTTGCAGTACTTACCCCTTGGCTAA	55
RESULT 7			
LOCUS B0650498			
DEFINITION B0650498 930 bp mRNA linear EST 15-JUL-2002			
ACCESSION AGENDOUT_8207703 NIH_MGC-100 Homo sapiens cDNA clone IMAGE:6285123			
VERSION B0650498.1 GI:2174670			
KEYWORDS 5', mRNA sequence.			
SOURCE EST.			
ORGANISM human.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE 1 (bases 1 to 930)			

AUTHORS	TITLE	JOURNAL	COMMENT
NIH-MGC	http://mgc.ncl.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)	Contact: Robert Strausberg, Ph.D. Email: cgapds-remail.nih.gov Tissue Procurement: CGAP (Stanford) CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1CM2478 row: 1 column: 20 High quality sequence stop: 636.
FEATURES	SOURCE	LOCATION/Qualifiers	
	1. .930	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_1 IMAGE:6283123" /clone_11b="NIH_MGC_100" /tissue_type="hepatocellular carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: Liver; Vector: pOT7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCGCAGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC library."	
BASE COUNT	ORIGIN		
Query Match	Best Local Similarity	Score 860.4; DB 14; Length 930;	
Matches 899; Conservative	97.0%; Pred. No. 2e-178;	Indels 6; Gaps 2	
	0; Mismatches 22;		
OY	301 CTCTTCCTGACTGCAGGCCCAATTACAGCAAAGCCGTGAAGCGCACCTGTG 360		
Db	1 CTCCTTCCTGACTGCAGGCCCAATTACAGCAAAGCCGTGAAGCGCACCTGTG 60		
OY	361 CAAGCAGCAGGGGGAGAGACCATTTCTTGTTGACTATGTAAGGGGTGTGGCAAGCCCTTCAT 420		
Db	61 CAAGCAGCAGGGGGAGAGACCATTTCTTGTTGACTATGTAAGGGGTGTGGCAAGCCCTTCAT 120		
OY	421 CAGGAGCTACCATTCTGAAGCCGCCACATTTCTGCTGACACAGAGAAAAGCCGTTTGTG 480		
Db	121 CAGGAGCTACCATTCTGAGCGGCACATTTCTGCTGCTGACACAGAGAAAAGCCGTTTGTG 180		
OY	481 TGCAGCCATGCGTGTGATCAAAAAATTCAACAACAAATCTTGAAGAAGAACTTTTGA 540		
Db	181 TGCAGCCAAATGCGTGTGATCAAAAAATTCAACAACAAATCTTGAAGAAGAACTTTTGA 240		
OY	541 ACGCAACATGAAATCAACAACAAATATATATATGACGTTTTGAGACTGTAGAAGAC 600		
Db	241 ACGCAACATGAAATCAACAACAAATATATATGACGTTTTGAGACTGTAGAAGAAC 300		
OY	601 CTTTAAAGAAATCAGCAGCTGAAAAATCCATAGTGCAGCATACCAATCAACCTCATTT 660		
Db	301 CTTTAAAGAAATCAGCAGCTGAAAAATCCATAGTGCAGCATACCAATCAACCTCATTT 360		
OY	661 CAAAGTACCCAGGAAGAGATGTGGGAAGACCTTTGCATACCCAGCAAGCTGAAAGCACA 720		
Db	361 CAAAGTACCCAGGAAGAGATGTGGGAAGACCTTTGCATACCCAGCAAGCTGAAAGCACA 420		
OY	721 TGCCAAGGCCCAAGAGGCTATGTATGTCAAAAAAGATGTTCCTTTGTGGCAAAAACATG 780		
Db	421 TGCCAAGGCCCAAGAGGCTATGTATGTCAAAAAAGATGTTCCTTTGTGGCAAAAACATG 480		
OY	781 GAGCGAAGCTTCTGAACATGTGAGACAAAACCATTAAGAGAAATACTATGTGAAGTATG 840		
Db	481 GAGCGAAGCTTCTGAACATGTGAGACAAAACCATTAAGAGAAATACTATGTGAAGTATG 540		

OY 841 CCGGAAACATTTAAAGCAAGATTACCTTAAGCAACACATGATAAACTCATGCCCGACA 900
|||||
Db 541 CCGGAAACATTTAAAGCAAGATTACCTTAAGCAACACATGATAAACTCATGCCCGACA 600
OY 901 AAGGATGTATGTCGCTGTCCAGAGAGGCTGTGAGAACTTACTACTGTGTTTAA 960
|||||
Db 601 AAGGATGTATGTCGCTGTCCAGAGAGGCTGTGAGAACTTACTACTGTGTTTAA 660
OY 961 TGTCCAAAGCCATATCTCTCTCTCCATGAGAAAGCCGCCCTTTGTGTGTAACATGC 1020
|||||
Db 661 TGTCCAAAGCCATATCTCTCTCTCCATGAGAAAGCCGCCCTTTGTGTGTAACATGC 720
OY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTAGGATGCTGTGTGATGATGA 1080
|||||
Db 721 TGGCTGTGGCAAAACATTTGCAATGAACAAGCTCTAGGATGCTGTGTGATGATGA 780
OY 1081 TECTGACAGAGAAATGAAAGCTCAAAAGTCAAAAATCTCGTGAAGAAACGAGTTTGGC 1140
|||||
Db 781 TECTGACAGAGAAATGAAAGCTCAAAAGTCAAAAATCTCGTGAAGAAACGAGTTTGGC 840
OY 1141 CTCTCATCTCACTGG--ATATATCCCTCCCAAAAGCAAGGCAAGG--CTTATC 1194
|||||
Db 841 CTCTCATCTCACTGGATATATATCCCTCCCAAAAGCAAGGCAAGGCTATCTCT 900
OY 1195 TTTGTGTCAAAACGAGAGTCAACCA 1221
|||
Db 901 TTTGTGTCAAAACGAGAGTCAACCA 927

RESULT 8
AL527493 957 bp mRNA linear EST 13-FEB-2001
LOCUS AL527493 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC024YE05 5
DEFINITION prime, mRNA sequence.
ACCESSION AL527493.1 GI:12790986
VERSION AL527493.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
Location/Qualifiers
1..957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC024YE05"
/clone_1id="LTI_NFL003_NBC3"
/sex="male"
/tissue-type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : filangel@life.com, URL : http://fulllength.invitrogen.com 1 others

BASE COUNT 283 a 240 c 227 g 206 t
Query Match 66.7%; Score 849.4; DB 9; Length 957;

Best Local Similarity 91.9%; Pred. No. 5.2e-176;
Matches 948; Conservative 1; Mismatches 7; Indels 75; Gaps 2;
OY 115 CCCGACGCTGTCTGGGACGCTGGCAGCGGCGCTTGCGCTGGGCTTGAAGCGCGGC 174
|||||
Db 1 CCCGACGCTGTCTGGGACGCTGGCAGCGGCGCTTGCGCTGGGCTTGAAGCGCGGC 60
OY 175 CCGGATCCCG 234
|||||
Db 61 CCGGATCCCG 120
OY 235 CATT--GCAGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCGCGCGCGCGCTTCCAGAGGT 293
|||||
Db 121 CATTAGCACCGCGGAGAGAGCTCAGCTCCGACCCCGCGCGCGCGCGCGCGCTTCCAGAGGT 180
OY 294 TCATGTCT 353
|||||
Db 181 TCATGTCT 240
OY 354 ACCGTGCAAGCAGCAGCGGCGGAGAGACCATTTGTGTGACTATGAAAGGTGTGCAAGC 413
|||||
Db 241 ACCGTGCAAGCAGCAGCGGCGGAGAGACCATTTGTGTGACTATGAAAGGTGTGCAAGG 300
OY 414 CCTTCATCGGCGGAGTACATCTGAGCGCGCAATTCGATCTCAGAGGAGAAAGCGGT 473
|||||
Db 301 CCTTCATCGGCGGAGTACATCTGAGCGCGCAATTCGATCTCAGAGGAGAAAGCGGT 360
OY 474 TTTGTGTGACGACGCTGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAC 533
|||||
Db 361 TTTGTGTGACGACGCTGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAC 420
OY 534 ATTTTGAACGCAAAATCATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAAT 593
|||||
Db 421 ATTTTGAACGCAAAATCATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAAT 480
OY 594 ACAAGACCTTTAAGAAGATGAGAGCTGAAATTCATCAGTCCAGCATACCAATGAAC 653
|||||
Db 481 ACAAGACCTTTAAGAAGATGAGAGCTGAAATTCATCAGTCCAGCATACCAATGAAC 540
OY 654 CTCTATTCAGGCTACCCAGAGAGAGTGTGGAAACCTTTGGATCACCCAGCAAGCTGA 713
|||||
Db 541 CTCTATTCAGGCTACCCAGAGAGAGTGTGGAAACCTTTGGATCACCCAGCAAGCTGA 550
OY 714 AACGACATGCCAAGGCCGACGAGAGGCTATGATGCTCAAAAGAGTGTCTTGTGGCAA 773
|||||
Db 551 AACGACATGCCAAGGCCGACGAGAGGCTATGATGCTCAAAAGAGTGTCTTGTGGCAA 586
OY 774 AACATGAGAGGAGCTGTGAATCATGTGAGAGAAACCATTAAGAGAAATCTATGTCG 833
|||||
Db 587 AACATGAGAGGAGCTGTGAATCATGTGAGAGAAACCATTAAGAGAAATCTATGTCG 646
OY 834 AAGTATCCCGGAAACATTTTAAAGCAAGATTAAGTCAAGCAACATGAAATCTATG 893
|||||
Db 647 AAGTATCCCGGAAACATTTTAAAGCAAGATTAAGTCAAGCAACATGAAATCTATG 706
OY 894 CCCGAGAAAGGATGATGCTGCTGTCCAGAGAGGCTGTGAGAAAGCACTTACTACTG 953
|||||
Db 707 CCCGAGAAAGGATGATGCTGCTGTCCAGAGAGGCTGTGAGAAAGCACTTACTACTG 766
OY 954 TGTTTAATTCCTCAAAAGCAATTCCTCTCTTCATGAGAGAAAGCCGCTTTGTGTGTG 1013
|||||
Db 767 TGTTTAATTCCTCAAAAGCAATTCCTCTCTTCATGAGAGAAAGCCGCTTTGTGTGTG 826
OY 1014 AACATCTGTGCTGTGGCAAAACATTTTGAATGAAGAAAGTCTCAGTACGATGCTGTG 1073
|||||
Db 827 AACATCTGTGCTGTGGCAAAACATTTTGAATGAAGAAAGTCTCAGTACGATGCTGTG 886
OY 1074 TACATATCTCTGACAGAGAAATGAAGCTCAAAAGTCAAAATCTCGTGAAGAAACGCA 1133
|||||
Db 887 TACATATCTCTGACAGAGAAATGAAGCTCAAAAGTCAAAATCTCGTGAAGAAACGCG 946
OY 1134 GTTTGGCTCT 1144
|||||

DB 947 TTGGGCTCT 957

RESULT 9
AL574698/c 906 bp mRNA Linear EST 16-FEB-2001

LOCUS AL574698 LTI_NFL006_P12 Homo sapiens cDNA clone CS0D1065123 3

DEFINITION prime, mRNA sequence.

ACCESSION AL574698

VERSION AL574698.1 GI:12935148

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 906)
Li, W.-B., Gruber, C., Jesse, J., and Polayes, D.
Full-length cDNA libraries and normalization
unpublished (2001)

JOURNAL Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr. Web : www.genoscope.cns.fr.

FEATURES
source
1..906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1065123"
/clone_1b="LTI_NFL006_P12"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five primers and enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 196 a 173 c 199 g 309 t 29 others

ORIGIN

Query Match 66.38; Score 844.6; DB 9; Length 906;
Best Local Similarity 95.18; Pred. No. 5.8e-175;
Matches 840; Conservative 26; Mismatches 17; Indels 0; Gaps 0;

391 TGACATGAAGGCTGTGGCAAGCCCTTCATCAGGACTACCAATCTGAGCGGCACATTTCT 450
|:|||||
906 TACTATGAGCGGTGTGGCAAGCCCTTCATCAGGACTACCAATCTGAGCGGCACATTTCT 847
|:|||||
451 GACTACACAGGAGAAAAGCCGTTTGTGTGCAAGCCACTGGCTGTGATCAAAATTTAA 510
|:|||||
846 GACTACACAGGAGAAAAGCCGTTTGTGTGCAAGCCACTGGCTGTGATCAAAATTTAA 787
|:|||||
511 CACAATAATCAAACTTGAAGAACTTTGAAGCAACATGAAATCAACAATAATTA 570
|:|||||
786 CACAATAATCAAACTTGAAGAACTTTGAAGCAACATGAAATSAACAATAATTA 727
|:|||||
571 TATATGCAAGTTTGAAGACTTAAAGAACCTTTAAGAAACATCAGACCTGAAATTTCA 630
|:|||||
726 TATATGCAAGTTTGAAGACTTAAAGAACCTTTAAGAAACATCAGACCTGAAATTTCA 667
|:|||||
631 TCAGTGCAGCATACCAATGAACTTATTCAGTGTACCCAGGAAGATGTGGGAACA 690
|:|||||
666 TCAGTGCAGCATACCAATGAACTTATTCAGTGTACCCAGGAAGATGTGGGAACA 607
|:|||||
691 CTTTGCATACCCAGCAAGCTGAAAGACATGCCAAGGCCACAGAGGCTATGTATTTCA 750
|:|||||
606 CTTTGCATACCCAGCAAGCTGAAAGACATGCCAAGGCCACAGAGGCTATGTATTTCA 547
|:|||||
751 AAAGAGATTTCTTTTGGCAAAAACATGAGCAACTTCTGAACATGTGAGAGAAC 810
|:|||||

DB 546 AAAAGATGTTCCCTTTTGGCAAAAACATGAGACGAACCTTCGAAACATGTGAGAGAAC 487

811 CCATTAAGAGGAAATTAATATGTAGTGAAGTATGCGGAAACATTTAAACGAAATTAATCT 870
|:|||||
486 CCATTAAGAGGAAATTAATATGTAGTGAAGTATGCGGAAACATTTAAACGAAATTAATCT 427
|:|||||
871 TAAGCAACATGAAACCTCATGCGCCAGAAAGAGATGATGCGCTGTGCCAAGGAAG 930
|:|||||
426 TAAGCAACATGAAACCTCATGCGCCAGAAAGAGATGATGCGCTGTGCCAAGGAAG 367
|:|||||
931 CTGTGGAGAACCTATATCTGTGTTTAAATCTCCAAAGCAATATCTCTCTCCATGCA 990
|:|||||
366 CTGTGGAGAACCTATATCTGTGTTTAAATCTCCAAAGCAATATCTCTCTCCATGCA 307
|:|||||
991 GGAAGCCGCGCTTTGTGTGTGAACATGCGCTGTGCCAAGCAATTTGAAGGAACA 1050
|:|||||
306 GGAAGCCGCGCTTTGTGTGTGAACATGCGCTGTGCCAAGCAATTTGAAGGAACA 247
|:|||||
1051 AAGTCTACATGAGCATGCTGTTGTACATGATCTCGACAGAAGAAATGAAAGCTCAAGT 1110
|:|||||
246 AATCTCAGTGTGCAAGGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 187
|:|||||
1111 CAAAAATCTGTGTGAAGGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 1170
|:|||||
186 YAAAAATCTGTGTGAAGGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 127
|:|||||
1171 AAGGAACAGGAGGAGGCTTATCTTTGTGTCAAAACGAGAGTCAACCACTGTGTGA 1230
|:|||||
126 AAGGAACAGGAGGAGGCTTATCTTTGTGTCAAAACGAGAGTCAACCACTGTGTGA 67
|:|||||
1231 ACACAAGATGCTGTGACAGTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 1273
|:|||||
66 ACACAAGATGCTGTGACAGTGTGCTGATGATGATGATGATGATGATGATGATGATGAT 24
|:|||||

RESULT 10
BM541851 1111 bp mRNA Linear EST 20-FEB-2002

LOCUS BM541851

DEFINITION AGNCOURT 6439143 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5531503

ACCESSION BM541851

VERSION BM541851.1 GI:18770849

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 1111)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-f@mail.nih.gov
Tissue Procurement: ATCC
DNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMLU)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMLU at:
http://image.llnl.gov
Plate: LAM12213 row: e column: 08
High quality sequence stop: 654.
Location/Qualifiers
1..1111
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5531503"
/clone_1b="NIH_MGC_71"
/tissue_type="telomerase-resistant"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: oligo dT.
Average insert size 2.1 kb.

BASE COUNT	287 a	313 c	295 g	216 t	
ORIGIN					
Query Match	66.3%; Score 844; DB 13; Length 1111;				
Best Local Similarity	92.9%; Pred. No. 7,9e-175;				
Matches 920; Conservative	0; Mismatches 60; Indels 10; Gaps 3				
QY	34	CTGTGACCGCGCGCCTCCCGAAGTGTGGCGGCGCGCGCGAAGGTTAGCGAGGGAGC	93		
Db	2	CTGTGTACCGCGCGCCTCCCGAAGTGTGGCGGCGCGCGAAGGTTAGCGAGGGAGC	61		
OY	94	CGTGGCGCGCGCGCGCGGTTCCCGGCAAGTGTCTGCGGACGTGGACGCCGTGGGCC	153		
Db	62	CGTGGCGCGCGCGCGCGGTTCCCGGCAAGTGTCTGCGGACGTGGACGCCGTGGGCC	121		
QY	154	TGGGCTTGGAGCGCGCGCGCGCGCGTGGATCCCGCGCGCGTGTGCGCGAGTGTGTGTG	213		
Db	122	TGGGCTTGGAGCGCGCGCGCGCGCGTGGATCCCGCGCGCGTGTGCGCGAGTGTGTGTG	181		
OY	214	CTTGACCATGTGCGGACGCGCTTATTTGACCGCGGAGAGCTCAGTCCGACCCCGCGCG	273		
Db	182	CTTGACCATGTGCGGACGCGCTTATTTGACCGCGGAGAGCTCAGTCCGACCCCGCGCG	241		
OY	274	CCCCGCGCTTCCGAGAGGTTATCTGTCTCTTCCCTGACATGCAAGCGCAATTACAGCAA	333		
Db	242	CCCCGCGCTTCCGAGAGGTTATCTGTCTCTTCCCTGACATGCAAGCGCAATTACAGCAA	301		
QY	334	AGCCTGAACTTTGACGCGCCTGTGCAAGCACACGGGGGAGAGACATTTGTTTGA	393		
Db	302	AGCCTGAACTTTGACGCGCCTGTGCAAGCACACGGGGGAGAGACATTTGTTTGA	361		
OY	394	CTTAAAGAGGTGTGCGCAAGGCTTCTATTCAGGACCTACATGTGAGCGGCCACATTTGAC	453		
Db	362	CTTAAAGAGGTGTGCGCAAGGCTTCTATTCAGGACCTACATGTGAGCGGCCACATTTGAC	421		
QY	454	TCACACAGGAGAAAGACCGGTTGTGTTGTGACACCACTGGCTGTATCAAAATTCACAC	513		
Db	422	TCACACAGGAGAAAGACCGGTTGTGTTGTGACACCACTGGCTGTATCAAAATTCACAC	481		
OY	514	AAATCAACTTAAGAAACATTTTGAACGCCAATATGAAATCAACAAAACAATATAT	573		
Db	482	AAATCAACTTAAGAAACATTTTGAACGCCAATATGAAATCAACAAAACAATATAT	541		
OY	574	ATGCAGTTTGAAGACTGTAGAAGACCTTTAAGAAACATGACGAGTGAATTCATCA	633		
Db	542	ATGCAGTTTGAAGACTGTAGAAGACCTTTAAGAAACATGACGAGTGAATTCATCA	601		
OY	634	GTCGACCATATACCAATGAACCTCTATTTCAAGTGTACCGAGGAAGATGTGGAAACACTT	693		
Db	602	GTCGACCATATACCAATGAACCTCTATTTCAAGTGTACCGAGGAAGATGTGGAAACACTT	661		
OY	694	TGCACTACCCAGCAAGCTGAACGACATGCCAAGGCCACGACGGCTATGTATGTCAAA	753		
Db	662	TGCACTACCCAGCAAGCTGAACGACATGCCAAGGCCACGACGGCTATGTATGTCAAA	721		
OY	754	AGGATGTTCTTGTGTGGCAAAAACATGAGCGGAACCTTCTGAACATGTGATGAGAAACCA	813		
Db	722	AGGATGTTCTTGTGTGGCAAAAACATGAGCGGAACCTTCTGAACATGTGATGAGAAACCC	781		
OY	814	TAAAGAGGAATCTATGTGSAAGTATGCCGGAAACATTTAAGCGAAAGATTACCTTAA	873		
Db	782	TAAAGAGGAATCTATGTGSAAGTATGCCGGGAACATTTAAGCGAAAGATTACCTTAA	841		
OY	874	GCAACACATATAATCTATGCCCCAGAAAGGATGTATGTGCGC---TGTCACAGAGAG	929		
Db	842	CCAACCCATGAAATCTATGCCCCAGAAAGGATGTATGTGCGC---TGTCACAGAGAG	901		
OY	930	GCCTGGAGAAAC---TATACTAGTGTGTTTAAATCTCAAAAGCCATAT---CCTGTCT	983		
Db	902	GCCTGGAGAAAC---TATACTAGTGTGTTTAAATCTCAAAAGCCATAT---CCTGTCT	961		
OY	984	TCCATGAGAAAGCCCGCTTTTGTGTGTG	1013		

Db	962	CCCTGGAGGAAAGACCGCCCTTTGGGGG	991
RESULT 11			
LOCUS	A0117724		
DEFINITION	A0117724 HEMBA1 Homo sapiens cDNA clone HEMBA1002055 5', mRNA sequence.	893 bp	linear EST 01-AUG-2002
ACCESSION	A0117724		
VERSION	A0117724.1	GI:10932715	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 893) Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isozaki,T.		
TITLE	HRI human cDNA project		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Takao Isozaki Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1..893 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="HEMBA1002055" /clone_lib="HEMBA1" /tissue_type="whole embryo, mainly head" /dev_stage="embryo, 10 weeks" /note="Vector: pME18SFL3"		
FEATURES			
source			
BASE COUNT	250 a 232 c 230 g 177 t	4 others	
ORIGIN			
Query Match	66.1%: Score 841.4; DB 9; Length 893;		
Best Local Similarity	99.5%: Pred. No. 2.9e-174;		
Matches 842: Conservative	0; Mismatches 4; Indels 0; Gaps 0;		
QY	60 TGTCCCGGCTCCGCGGAAGTTACAGACAGAGCGTGGGCGCGCGCTTCGCGG	119	
DB	1 TGTCCGCGCTCCGCGGAAGTTACAGACAGAGCGTGGGCGCGCGCTTCGCGG	60	
QY	120 CACGTCTCTGGGACGTGGCAGCGCGGCTTGAGGCTTGAGAGGCGCGCCCTGG	179	
DB	61 CACGTCTCTGGGACGTGGCAGCGCGGCTTGAGGCTTGAGAGGCGCGCCCTGG	120	
QY	180 ATTCGCGCGGCGTGTGTGCGCAGTGTGTCTCTTACCATCGCCGACGCTTCATTG	239	
DB	121 ATTCGCGCGGCGTGTGTGCGCAGTGTGTCTCTTACCATCGCCGACGCTTCATTG	180	
QY	240 CAGCGGCGGACGACTCAGCTCCGACCCCGCGCGCCCGGCGCTTCCAGAGGTTTCATCT	299	
DB	181 CAGCGGCGGACGACTCAGCTCCGACCCCGCGCGCCCGGCGCTTCCAGAGGTTTCATCT	240	
QY	300 GCTCTTCCCTGACTCGAGCGGCATTTACAGCAAGCCTTGAGACCTTGACGCGCACCTGT	359	
DB	241 GCTCTTCCCTGACTCGAGCGGCATTTACAGCAAGCCTTGAGACCTTGACGCGCACCTGT	300	
QY	360 GCAGGACACAGGGGGGAGAGCAATTTGTTGGAGCTATGAAGGGGTGGCAAGCCCTTCA	419	
DB	301 GCAGGACACAGGGGGGAGAGCAATTTGTTGGAGCTATGAAGGGGTGGCAAGCCCTTCA	360	
QY	420 TCAGGACTACATCTGAGCGCGCACATTTCTGACTCACACAGGAGAAACCCGTTTGTT	479	

```

|||||
Db 361 TCAGGAGCTACCTGCGCGCCACATTCCTGACACACAGAGAAAAGCCGTTGTTT 420
Oy 480 GTGACGCCACTGGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAGAAACATTTTG 539
Db 421 GTGACGCCAAATGGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAGAAACATTTTG 480
Oy 540 AACGCAAAATGAAAATTCACAAAACAAATATATATGCGATTTCAGCTGTAGAGAGA 599
Db 481 AACGCAAAATGAAAATTCACAAAACAAATATATATGCGATTTCAGCTGTAGAGAGA 540
Oy 600 CCTTTAAGAAACATCAGCAGCTGAAATTCATGATGCGCAGCATACCAATGAACCTCTAT 659
Db 541 CCTTTAAGAAACATCAGCAGCTGAAATTCATGATGCGCAGCATACCAATGAACCTCTAT 600
Oy 660 TCAGCTGATCCAGGAGAGATGGGAAACACCTTTCATCACCGACGCAAGTGAAAGAGAC 719
Db 601 TCAGCTGATCCAGGAGAGATGGGAAACACCTTTCATCACCGACGCAAGTGAAAGAGAC 660
Oy 720 ATGCCAAGCCCGACAGGCGTATGTATGCAAAAAGGATGTCCTTTGTGGCAAAAACAT 779
Db 661 ATGCCAAGCCCGACAGGCGTATGTATGCAAAAAGGATGTCCTTTGTGGCAAAAACAT 720
Oy 780 GGACGGAACCTCTGAAACATGTGAGAGAACCCATAAAGGAAATACATGTGAGAT 839
Db 721 GGACGGAACCTCTGAAACATGTGAGAGAACCCATAAAGGAAATACATGTGAGAT 780
Oy 840 GCCGGAACCATTTAAACGCAACATTAACCTTAAGCAGACATGAAACCTCAGCCGAC 899
Db 781 GCCGGAACCATTTAAACGCAACATTAACCTTAAGCAGACATGAAACCTCAGCCGAC 840
Oy 900 AAAGG 905
Db 841 AAANG 846

```

```

RESULT 12
LOCUS B0644275 902 bp mRNA linear EST 15-JUL-2002
DEFINITION AGNCOURT 8355851 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6285421
5' mRNA sequence.
ACCESSION B0644275
VERSION B0644275.1 GI:21768447
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mgsa.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rgsb@nsl.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2484 row: 3 column: 14
High quality sequence stop: 642.
Location/Qualifiers
1. 902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pORF7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned

```

```

FEATURES
source

```

```

into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library.
BASE COUNT 301 a 208 c 197 g 194 t 2 others
ORIGIN
Query Match 65.7% Score 836; DB 14; Length 902;
Best Local Similarity 98.3%; Pred. No. 4,5e-173;
Matches 875; Conservative 0; Mismatches 12; Indels 3; Gaps 3;
Oy 300 GCTCCTCCCTGACGCGAGCGCCAAATTCAGCAAAACCTGGAAGCTTGACGGCAGCTGT 359
Db 11 GCTCCTCCCTGACGCGAGCGCCAAATTCAGCAAAACCTGGAAGCTTGACGGCAGCTGT 70
Oy 360 GCAGCAGCAGGAGGAGACACATTTGTTGTGACTATGAGGGGTGGCAAGGCTTCA 419
Db 71 GC-AACACACGAGGAGAGACATTTGTTGTGACTATGAGGGGTGGCAAGGCTTCA 129
Oy 420 TCAGGAGCTACCATCTGAGCCGCCACATCTGACTACACAGAGAAAAGCCGTTGTTT 479
Db 130 TCAGGAGCTACCATCTGAGCCGCCACATCTGACTACACAGAGAAAAGCCGTTGTTT 189
Oy 480 GTGACGCCACTGGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAGAAACATTTTG 539
Db 190 GTGACGCCAAATGGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAGAAACATTTTG 249
Oy 540 AACGCAAAATGAAAATTCACAAAACAAATATATATGCGATTTCAGCTGTAGAGAGA 599
Db 250 AACGCAAAATGAAAATTCACAAAACAAATATATATGCGATTTCAGCTGTAGAGAGA 309
Oy 600 CCTTTAAGAAACATCAGCAGCTGAAATTCATGATGCGCAGCATACCAATGAACCTCTAT 659
Db 310 CCTTTAAGAAACATCAGCAGCTGAAATTCATGATGCGCAGCATACCAATGAACCTCTAT 369
Oy 660 TCAGCTGATCCAGGAGAGATGGGAAACACCTTTCATCACCGACGCAAGTGAAAGAGAC 719
Db 370 TCAGCTGATCCAGGAGAGATGGGAAACACCTTTCATCACCGACGCAAGTGAAAGAGAC 429
Oy 720 ATGCCAAGCCCGACAGGCGTATGTATGCAAAAAGGATGTCCTTTGTGGCAAAAACAT 779
Db 430 ATGCCAAGCCCGACAGGCGTATGTATGCAAAAAGGATGTCCTTTGTGGCAAAAACAT 489
Oy 780 GGACGGAACCTCTGAAACATGTGAGAGAACCCATAAAGGAAATACATGTGAGAT 839
Db 490 GGACGGAACCTCTGAAACATGTGAGAGAACCCATAAAGGAAATACATGTGAGAT 549
Oy 840 GCCGGAACCATTTAAACGCAACATTAACCTTAAGCAGACATGAAACCTCAGCCGAC 899
Db 550 GCCGGAACCATTTAAACGCAACATTAACCTTAAGCAGACATGAAACCTCAGCCGAC 609
Oy 900 AAAGGATGTATGTGCGCTGTCACAGAGAGGCTGTGAAGAACCTTACTACTGTGTTT 959
Db 610 AAAGGATGTATGTGCGCTGTCACAGAGAGGCTGTGAAGAACCTTACTACTGTGTTT 669
Oy 960 ATCTCCAAAGCATATCTCTCTCTTCATGAGAAACCGCCCTTTTGTGTGAACATG 1019
Db 670 ATCTCCAAAGCATATCTCTCTCTTCATGAGAGAAACCGCCCTTTTGTGTGAACATG 729
Oy 1020 CTGGCTGTGGCAAAACATTTGCATGAAACAAAGTGTGATGAGCTGTGTGATCATG 1079
Db 730 CTGGCTGTGGCAAAACATTTGCATGAAACAAAGTGTGATGAGCTGTGTGATCATG 789
Oy 1080 ATCTGACAGAGAAATGAGAGTCAAAATCTGTAAGAAATCTGTAAGAAAGGATTTGG 1139
Db 790 ATCTGACAGAGAAATGAGAGTCAAAATCTGTAAGAAATCTGTAAGAAAGGATTTGG 848
Oy 1140 CCTTCATCTCAGTGCATATAT -CCTGCCCAAGGAAAGCAAGGCAAG 1188
Db 849 CCTTCATCTCAGTGCATATATCCCTGCCCAAGGAAAGCAAGGCGCAG 898

```

```

RESULT 13
LOCUS      AL578193      933 bp      mRNA      linear      EST 16-FEB-2001
DEFINITION AL578193 LTI_NFL006_P12 Homo sapiens cDNA clone CS0DK002YJ18 3
prime, mRNA sequence.
ACCESSION  AL578193
VERSION     AL578193.1  GI:12942034
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 933)
AUTHORS     Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK002YJ18"
/clone_id="LTI_NFL006_P12"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact: Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT  202 a 192 c 207 g 326 t 6 others
ORIGIN
Query Match 65.28; Score 830.2; DB 9; Length 933;
Best Local Similarity 96.58; Pred. No. 8.3e-172;
Matches 854; Conservative 4; Mismatches 7; Indels 2; Gaps 2;
408 CCAAGGCTTCATCAGGACTACCATCTGAGCCGCCACATTCGACTCACAGAGGAAA 467
|||||
933 GCAAGGCTTATCA-GGACTACGATCTGAGCCGCCACATTCMACTCACAGAGGAAA 875
468 ACCGCTTTGTTTGTGACCCACTGCGTGATCAAAAATTCACACAAATCAAACTTCA 527
|||||
874 ACCCGTTTGTGTGACCCCAATGCGTGATCAAAAATTCACACAAATCAAACTTCA 815
528 AGAAGCATTTGAAGCGCAACATGAAATCAACAAACATATATATCAGTTTGAAG 587
|||||
814 AAGAACATTTGAAGCGCAACATGAAATCAACAAACATATATATCAGTTTGAAG 755
588 ACTGTAAAGACCTTTAAGAAACATGACAGCTGAAATCCATCAGTCCAGCATPCCA 647
|||||
754 ACTGTAAAGACCTTTAAGAAACATGACAGCTGAAATCCATCAGTCCAGCATPCCA 695
648 ATGACCTCTATTCAGAGTACCCAGAGAGATGTGGAAACACTTGGATCACCAGACA 707
|||||
694 ATGACCTCTATTCAGAGTACCCAGAGAGATGTGGAAACACTTGGATCACCAGACA 635
708 AACTGAAGACATGCCAAGGCCACAGAGGCTATGATGTCAAAAAGATGTTCTTTTG 767
|||||
634 AACTGAAGACATGCCAAGGCCACAGAGGCTATGATGTCAAAAAGATGTTCTTTTG 575
768 TGGCAAAAACATGAGAGCAACTTGTGAACATGTGAGAGAAACCACTAAGAGAAATAC 827
|||||
574 TGGCAAAAACATGAGAGCAACTTGTGAACATGTGAGAGAAACCACTAAGAGAAATAC 515

```

```

QY 828 TATGTGAAGTATGCCGGAAGACATTTAAACGCCAAGATTCCTTAAGCACATGAAA 887
|||||
DB 514 TATGTGAAGTATGCCGGAAGACATTTAAACGCCAAGATTCCTTAAGCACATGAAA 455
888 CTCATGCCCCAGAAAAGGATGTATGTGCTGTCACAGAGAGGCTGTGAAGACCTATA 947
DB 454 CTCATGCCCCAGAAAAGGATGTATGTGCTGTCACAGAGAGGCTGTGAAGACCTATA 395
QY 948 CTACTGTGTTTAACTCCAAAGCCATATCTCTCTCTTCCATGAGCAAGCGCCCTTTTG 1007
|||||
DB 394 CAACTGTGTTTAACTCCAAAGCCATATCTCTCTCTTCCATGAGCAAGCGCCCTTTTG 335
1008 TGTGTGAACATGCGTGGCTGTGGCAAAACATTTGCATGAAACAAAGTCCACTAGCATG 1067
|||||
DB 334 TGTGTGAACATGCGTGGCTGTGGCAAAACATTTGCATGAAACAAAGTCCACTAGCATG 275
QY 1068 CTGTTGTACATGATCTTCGACAGAGAAATGAAGCTCAAGTCAAAATCTCTGAAA 1127
|||||
DB 274 CTGTTGTACATGATCTTCGACAGAGAAATGAAGCTCAAAATCTCTGAAA 215
QY 1128 AACGAGTTTGGCTCTCATCTCAGTGAATATCCCTCCCAAAAGAAACAAAGGCAAG 1187
|||||
DB 214 AACGAGTTTGGCTCTCATCTCAGTGAATATCCCTCCCAAAAGAAACAAAGGCAAG 155
QY 1188 GCTTATCTTTGTGTCAAAAGGAGAGTCAACCACTGTGTGGAAGCAAGATGCTCTGA 1247
|||||
DB 154 GCTTATCTTTGTGTGTCAAAAGGAGAGTCAACCACTGTGTGGAAGCAAGATGCTCTGA 95
QY 1248 CAGTTGCA-GTACTTACCTTGCTTAA 1273
|||||
DB 94 CAGTTGCAAGCACTTACCTTGCTTAA 68

```

```

RESULT 14
LOCUS      B0278693      1089 bp      mRNA      linear      EST 07-MAY-2002
DEFINITION AGENCOURT_7049694 NIH_MGC_107 Homo sapiens cDNA clone IMAGE:5805333
5', mRNA sequence.
ACCESSION  B0278693
VERSION     B0278693.1  GI:20488901
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1089)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
Email: cagpds@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.lnl.gov
Plate: L1CM2044 row: f column: 22
High quality sequence stop: 610.
FEATURES
source
1. 1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5805333"
/clone_id="NIH_MGC_107"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Library constructed by

```

Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library.

BASE COUNT 296 a 289 c 283 g 221 t
ORIGIN

Query Match 65.2% Score 830; DB 14; Length 1089;
Best Local Similarity 94.6%; Pred. No. 9.3e-172;
Matches 927; Conservative 0; Mismatches 40; Indels 13; Gaps 6;

70 TCGCCGGAAGGTTACAGCAGGAGCGCTGGCGCGCGCGCTTCCGGCAGCTGCTC 129
|||||
1 TCGCCGGAAGGTTACAGCAGGAGCGCTGGCGCGCGCGCTTCCGGCAGCTGCTC 60
|||||
130 GGCAGCTGAGCGCGCTGGCGCTTGGAGGCGCGCGCGCTTGGAGCGCGCGC 189
|||||
61 GGCAGCTGAGCGCGCTGGCGCTTGGAGGCGCGCGCGCTTGGAGCGCGCGC 120
|||||
190 GCTGTCGCCGAGTGGCTGCTGCTTGGACCATGCGCGAGCGCTTATGACCGCGCA 249
|||||
121 CCGTGTCCCGCAGTGGCTGCTGCTTGGACCATGCGCGAGCGCTTATGACCGCGCA 180
|||||
250 GAGCTCAGCTCGAGCGCGCGCGCGCGCGCTTCCAGAGGTTATGCTGCTTCCG 309
|||||
181 GAGCTCAGCTCGAGCGCGCGCGCGCGCGCTTCCAGAGGTTATGCTGCTTCCG 240
|||||
310 TCAGTCGACGCGCAATTAACAGAAAGCTTGAAGCGCGACCTGTCAGCAGCAGC 369
|||||
241 TCAGTCGACGCGCAATTAACAGAAAGCTTGAAGCGCGACCTGTCAGCAGCAGC 300
|||||
370 GGGGAGAGACCATTTGTTGTTGACTATGAAGGTTGCGAAGCGCTTATCAGGAGCTA 429
|||||
301 GGGGAGAGACCATTTGTTGTTGACTATGAAGGTTGCGAAGCGCTTATCAGGAGCTA 360
|||||
430 CCATGTGAGCGCGCAATTTGACTATGACACAGAGAAAGCGCTTGTGTTGTCAGCCAC 489
|||||
361 CCATGTGAGCGCGCAATTTGACTATGACACAGAGAAAGCGCTTGTGTTGTCAGCCAC 420
|||||
490 TGGCTGTGATCAAAAATTCACACAAATCAAACTGAGAAACATTTTGAAGCGCAACA 549
|||||
421 TGGCTGTGATCAAAAATTCACACAAATCAAACTGAGAAACATTTTGAAGCGCAACA 480
|||||
550 TGAATAATCAACAAAACATATATATGCACTTTTGAAGAGCTGAGACACTTTAAAGA 609
|||||
481 TGAATAATCAACAAAACATATATATGCACTTTTGAAGAGCTGAGACACTTTAAAGA 540
|||||
610 ACATGAGCAGCTGAAAATCATCATGTCAGCATACCATGAACTCTATTCAAGCTTAC 669
|||||
541 ACATGAGCAGCTGAAAATCATCATGTCAGCATACCATGAACTCTATTCAAGCTTAC 600
|||||
670 CAGGAAGGATGTGGGAACACTTGGATCAGCCAGCAAGCTGAAACGATGCCAAGCC 729
|||||
601 CAGGAAGGATGTGGGAACACTTGGATCAGCCAGCAAGCTGAAACGATGCCAAGCC 660
|||||
730 CCAGCAGGCTATGATGTCAAAAGAGATGTTCCCTTTGGGCAAAAACATGAGCAGCACT 789
|||||
661 CCAGCAGGCTATGATGTCAAAAGAGATGTTCCCTTTGGGCAAAAACATGAGCAGCACT 720
|||||
790 TCGTGAACATGTGAGAGAAACCCATAAGAGAAATATATATGTAAGATGCGCGAAGAAC 849
|||||
721 TCGTGAACATGTGAGAGAAACCCATAAGAGAAATATATATGTAAGATGCGCGAAGAAC 780
|||||
850 ATTAAAGCAAGATTAATTAAGCAACATGAAACTCAT--GCCCAAGAAAGGAT 907
|||||
781 ATTAAAGCAAGATTAATTAAGCAACATGAAACTCATGCGCGCGCAAGAAAGGAT 840
|||||
908 GATATGCC--TGTCGAAGAGAGAGCTGTGG--AAGAACTATATCTACTGT--GTTTAATC 962
|||||
841 GATATGCCCTGTGTCGAAGAGAGAGCTGTGGAAGAAACCTATTTCCACTGTGGGTTAAATCC 900
|||||
963 TGCAGAGCATATC---CTCTCTTCCATGAGCAAGCGCGCTTTT---GTTGTGAAC 1016
|||||

Db 901 TCCCAAGCATATCCCTCTCTCTTCCCTGAGAGAAAGCCCGCTTTTGGGGGGGAGAC 960
QY 1017 ATGCTGCTGTGGCAAAACA 1036
|||||
Db 961 ATGCTGCTGTGGCGCAAAA 980

RESULT 15
LOCUS BC830080 927 bp mRNA linear EST 22-MAY-2001
DEFINITION 602764826F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4906710 5',
mRNA sequence.
ACCESSION BC830080
VERSION BC830080.1 GI:14177667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 927)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gsapbs-remail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LCM1810 row: d column: 07
High quality sequence stop: 880.

FEATURES

source 1..927
location/Qualifiers
1..927
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4906710"
/clone_1lb="NIH_MGC_42"
/tissue_type="epithelioid carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; site_1: XhoI;
site_2: EcoRI; cDNA made by 01.90-07 priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(6). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library. |"

BASE COUNT

264 a 244 c 229 g 190 t

ORIGIN

Query Match 65.0% Score 827.4; DB 12; Length 927;
Best Local Similarity 98.6%; Pred. No. 3.4e-171;
Matches 908; Conservative 0; Mismatches 6; Indels 7; Gaps 7;

98 GCGCGGGCGCGCGCTTCCCGCACGTCGTCGGCACGTGCGAGCGCGCTGGCCTGG 157
|||||
Db 1 GCGCGGGCGCGCGCTTCCCGCACGTCGTCGGCACGTGCGAGCGCGCGCTGGCCTGG 60
|||||
QY 158 CTGGAGAGCGCGCGCGCTGATCCCGCGCGCGGTGCGCGCGAGTCGGTGTGCTTTG 217
|||||
Db 61 CTGGAGAGCGCGCGCGCTGATCCCGCGCGCGGTGCGCGCGAGTCGGTGTGCTTTG 120
|||||
QY 218 ACCATCGCGAGCGGCTTATTCAGCGCGCGGAGAGCTGACGTCGACCGCGCGCGCC 277
|||||
Db 121 ACCATCGCGAGCGGCTTATTCAGCGCGCGGAGAGCTGACGTCGACCGCGCGCGCC 180
|||||
QY 278 GCGCTTCCAGAGAGGTTATCTGCTCTCTCTGACTGAGCGCAATTAAGCAAGCC 337
|||||
Db 181 GCGCTTCCAGAGAGGTTATCTGCTCTCTCTGACTGAGCGCAATTAAGCAAGCC 240
|||||

```
OY 338 TGGAGCTTGGACGGCCACCTGTGCAAGCACACGGGGAGAGACCATTGTTGTGACTAT 397
    |||||||
Db 241 TGGAAAGCTTGGACGGCCACCTGTGCAAGCACACGGGGAGAGACCATTGTTGTGACTAT 300
OY 398 GAAGGCTGTGGCAAGGCCCTTCATCAGGACTACCATCTGAGCCGCACATCTGACTCAC 457
    |||||||
Db 301 GAAGGCTGTGGCAAGGCCCTTCATCAGGACTACCATCTGAGCCGCACATCTGACTCAC 360
OY 458 ACAGAGAGAAAAAGCCGTTGTTGTGTGCAGCCATGCGCTGTGATCAAAAATTCAACACAAA 517
    |||||||
Db 361 ACAGAGAGAAAAAGCCGTTGTTGTGTGCAGCCATGCGCTGTGATCAAAAATTCAACACAAA 420
OY 518 TCAAACTTGAAGAAACATTTTGAACGCAAAACATGAAATCAACAAACATATATATATGC 577
    |||||||
Db 421 TCAAACTTGAAGAAACATTTTGAACGCAAAACATGAAATCAACAAACATATATATATGC 480
OY 578 AGTTTGAAGACTGTAGAGAGACCTTTAAGAAACATCAGAGCTGAAAAATCCATCAGTGC 637
    |||||||
Db 481 AGTTTGAAGACTGTAGAGAGACCTTTAAGAAACATCAGAGCTGAAAAATCCATCAGTGC 540
OY 638 CAGCATACCAATGAACCTCTATTCAGTGTACCCAGAGAGATGTGGAAGACACTTTGCA 697
    |||||||
Db 541 CAGCATACCAATGAACCTCTATTCAGTGTACCCAGAGAGATGTGGAAGACACTTTGCA 600
OY 698 TCACCCAGCAAGCTGAAGACATGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGGA 757
    |||||||
Db 601 TCACCCAGCAAGCTGAAGACATGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGGA 660
OY 758 TGTTCCTTGTGGCAAAAACATGAGACGAACTTCTGAAACATGTGAGAGAAACCCAT-AA 816
    |||||||
Db 661 TGTTCCTTGTGGCAAAAACATGAGACGAACTTCTGAAACATGTGAGAGAAACCCATAA 720
OY 817 AAGAGAAATACATATGTG-AACTATGCCGGAAAA-CAITTTAAAGCAAAAG-TTACCTTAA 873
    |||||||
Db 721 AAGAGAAATACATATGTGAAAGTATGCGGAAAAACCAITTTAAAGCAAAAGTTTACCTTAA 780
OY 874 GCACACATGAAAACCTCATGCCCCCAGAAAGGATGTATGCGCTGTCCAAGAGAAAGGCT- 932
    |||||||
Db 781 GCACACATGAAAACCTCATG-CCCAAGAGGATCTATGTGCTGTCCAAGAGAGGCTG 839
OY 933 GTGGAAGAACTATACCTGTGTGTTAAT-CTCCAAAGCATATCTCTCTTCCATGAG 991
    |||||||
Db 840 GTGGAAGAACTATACCTGTGTGTTAATCTCCAAAGCATATCTCTCTTCCATGAG 899
OY 992 GAAAGCCGCCCTTTTGTGTGT 1012
    |||||||
Db 900 GAAAGCCGCCCTTTTGTGTGT 920
```

Search completed: February 10, 2003, 17:44:07
Job time : 2015.41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:38:15 : Search time 68.105 Seconds
(without alignments) 8859.935 Million cell updates/sec

Title: US-09-831-426c-3

Perfect score: 1273

Sequence: 1 atgcgcagcagcgcgcgca.....cagctactaccctgctgctaa 1273

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 408267 seqs, 237001491 residues

Total number of hits satisfying chosen parameters: 816534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/PCRUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418.4	32.9	439	10	US-09-920-300A-326 Sequence 326, App
2	418.4	32.9	439	12	US-10-033-528-326 Sequence 326, App
3	416.8	33.7	439	10	US-09-920-300A-689 Sequence 689, App
4	416.8	33.7	439	12	US-10-033-528-689 Sequence 689, App
5	234	18.4	234	9	US-08-786-692-4064 Sequence 4064, App
6	234	18.4	234	10	US-09-920-300A-610 Sequence 610, App
7	234	18.4	234	12	US-10-033-528-610 Sequence 610, App
8	226	17.8	362	9	US-09-292-758-68 Sequence 68, Appl
9	152.4	12.0	947	10	US-09-995-973-2 Sequence 2, Appl
10	152.4	12.0	947	12	US-09-732-348-5 Sequence 5, Appl
11	152.4	12.0	995	10	US-09-995-973-1 Sequence 1, Appl
12	152.4	12.0	995	12	US-09-732-348-4 Sequence 4, Appl
13	141.6	11.1	449	10	US-08-783-590-5443 Sequence 5443, App
14	86.6	6.8	1111	10	US-09-833-381-1175 Sequence 1175, App
15	81	6.4	368	10	US-09-920-300A-1747 Sequence 1747, App
16	81	6.4	368	12	US-10-033-528-1747 Sequence 1747, App
17	80.4	6.3	2400	10	US-09-864-761-18874 Sequence 18874, App
18	69.2	5.4	2960	9	US-09-734-329-1 Sequence 1, Appl
19	64.6	5.1	292	9	US-09-942-090-14 Sequence 14, Appl

20	64.6	5.1	292	10	US-09-925-796-14 Sequence 14, Appl
21	64.6	5.1	292	10	US-09-941-450-14 Sequence 14, Appl
22	63	4.9	298	9	US-09-942-087A-14 Sequence 14, Appl
23	63	4.9	298	10	US-09-779-233-2 Sequence 2, Appl
24	63	4.9	298	10	US-09-844-508-10 Sequence 10, Appl
25	62	4.9	3213	12	US-10-002-600-76 Sequence 76, Appl
26	61.8	4.9	547	10	US-09-667-701-4943 Sequence 4943, App
27	61.8	4.9	3197	10	US-09-764-864-1647 Sequence 1647, App
28	60.6	4.8	589	9	US-09-942-087A-29 Sequence 29, Appl
29	60.6	4.8	589	10	US-09-779-233-17 Sequence 17, Appl
30	60.2	4.8	589	10	US-09-844-508-28 Sequence 28, Appl
31	60.2	4.7	900	10	US-09-764-864-211 Sequence 211, App
32	60.2	4.7	1602	10	US-09-833-790-417 Sequence 417, App
33	60	4.7	259	10	US-09-960-352-4925 Sequence 4925, App
34	60	4.7	882	10	US-09-764-864-642 Sequence 642, App
35	58.2	4.5	298	9	US-09-942-087A-16 Sequence 16, Appl
36	58.2	4.5	298	10	US-09-779-233-4 Sequence 4, Appl
37	58.2	4.6	298	10	US-09-844-508-17 Sequence 17, Appl
38	55.4	4.4	1301	9	US-09-954-531-950 Sequence 950, App
39	55.4	4.4	1301	10	US-09-880-107-1613 Sequence 1613, App
40	55.4	4.4	1301	10	US-09-967-768A-120 Sequence 120, App
41	55.4	4.4	2572	10	US-09-919-172-73 Sequence 17, App
42	55.4	4.4	3332	9	US-10-071-766-137 Sequence 137, App
43	54.8	4.3	3138	10	US-09-880-107-1716 Sequence 116, App
44	54.6	4.3	396	9	US-08-970-966-119 Sequence 119, App
45	54.6	4.3	396	10	US-09-825-294-119 Sequence 119, App

ALIGNMENTS

RESULT 1
US-09-920-300A-326/c
: Sequence 326, Application US/0920300A
: Patent No. US20020136728A1
: GENERAL INFORMATION:
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jiangchun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547
: CURRENT APPLICATION NUMBER: US/09/920.300A
: CURRENT FILING DATE: 2001-07-31
: NUMBER OF SEQ ID NOS: 1789
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 326
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-920-300A-326

Query Match 32.9%; Score 418.4; DB 10; Length 439;

Best Local Similarity 99.8%; Pred. No. 8.4e-103;

Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	854	AAACGCAAGATTACCTTAGCAGACACTGAAACACTAGTCCCGCAAGAGGATGATGT	913
DB	439	AAACGCAAGATTACCTTAGCAGACACTGAAACACTAGTCCCGCAAGAGGATGATGT	380
QY	914	CGCTGTCACAGAGAGCGCTGTCAGAGACATATCTGCTGTTATCTCCAAACCAT	973
DB	379	CGCTGTCACAGAGAGCGCTGTCAGAGACATATCTGCTGTTATCTCCAAACCAT	320
QY	974	ATTCCTCTCTTCATGAGAGAAAGCCGCTTTTGTTGTTGATGACATGCTGCTGCAAA	1033
DB	319	ATTCCTCTCTTCATGAGAGAAAGCCGCTTTTGTTGTTGATGACATGCTGCTGCAAA	260
QY	1034	ACATTTTGCAATGAACAAAGTCTCACTAGGCATCTCTTTACATGATCTGCACAGAG	1093
DB	259	ACATTTTGCAATGAACAAAGTCTCACTAGGCATCTCTTTGACATGATCTGCACAGAG	200

```
OY 1094 AAATGAACTGATGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCATCTCACT 1153
      |||||||
Db 199 AAATGAACTGATGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCATCTCACT 140
OY 1154 GGATATATATCCCTCCCAAAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1213
      |||||||
Db 139 GGATATATATCCCTCCCAAAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 80
OY 1214 TCACCCCACTGTGTGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCACT 1273
      |||||||
Db 79 TCACCCCACTGTGTGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCACT 20
```

RESULT 2

```
US-10-033-528-326/c
: Sequence 326, Application us/10033528
: Patent No. US20020131971A1
: GENERAL INFORMATION:
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jianshun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.547C1
: CURRENT APPLICATION NUMBER: US/10/033.528
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 326
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-033-528-326
```

```
Query Match 32.9%: Score 418.4; DB 12; Length 439;
Best Local Similarity 99.8%; Pred. No. 8.4e-103;
```

```
Matches 419; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 854 AAACGCAAGATTACCTTACGACACATGAAACCTCATGCCCCAGAAAGGATGTATGT 913
      |||||||
Db 439 AAACGCAAGATTACCTTACGACACATGAAACCTCATGCCCCAGAAAGGATGTATGT 380
OY 914 CGCTGTCCAAAGAGGCTGTGGAAGAACCTATCTACTGTGTTTAACTCCAAAGCCAT 973
      |||||||
Db 379 CGCTGTCCAAAGAGGCTGTGGAAGAACCTATCTACTGTGTTTAACTCCAAAGCCAT 320
OY 974 ATCTCTCTCCCTCATGAGGAAAGCGCCCTTTGTGTGAACATGCTGGCTGTGCAAA 1033
      |||||||
Db 319 ATCTCTCTCCCTCATGAGGAAAGCGCCCTTTGTGTGAACATGCTGGCTGTGCAAA 260
OY 1034 ACATTGGCAATGAACAAAGTCTCACTAGCATGCTCTGTACATGATCTGCACAAGAG 1093
      |||||||
Db 259 ACATTGGCAATGAACAAAGTCTCACTAGCATGCTCTGTACATGATCTGCACAAGAG 200
OY 1094 AAATGAACTGATGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCATCTCACT 1153
      |||||||
Db 199 AAATGAACTGATGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCATCTCACT 140
OY 1154 GGATATATATCCCTCCCAAAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1213
      |||||||
Db 139 GGATATATATCCCTCCCAAAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 80
OY 1214 TCACCCCACTGTGTGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCACT 1273
      |||||||
Db 79 TCACCCCACTGTGTGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCACT 20
```

RESULT 3

```
US-09-920-300A-689
: Sequence 689, Application us/09920300A
: Patent No. US200201316728A1
: GENERAL INFORMATION:
```

```
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jianshun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.547
: CURRENT APPLICATION NUMBER: US/09/920.300A
: NUMBER OF SEQ ID NOS: 1789
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 689
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-920-300A-689
```

```
Query Match 32.7%: Score 416.8; DB 10; Length 439;
Best Local Similarity 99.5%; Pred. No. 2.3e-102;
```

```
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY 854 AAACGCAAGATTACCTTACGACACATGAAACCTCATGCCCCAGAAAGGATGTATGT 913
      |||||||
Db 1 AAACGCAAGATTACCTTACGACACATGAAACCTCATGCCCCAGAAAGGATGTATGT 60
OY 914 CGCTGTCCAAAGAGGCTGTGGAAGAACCTATCTACTGTGTTTAACTCCAAAGCCAT 973
      |||||||
Db 61 CGCTGTCCAAAGAGGCTGTGGAAGAACCTATCTACTGTGTTTAACTCCAAAGCCAT 120
OY 974 ATCTCTCTCCCTCATGAGGAAAGCGCCCTTTGTGTGAACATGCTGGCTGTGCAAA 1033
      |||||||
Db 121 ATCTCTCTCCCTCATGAGGAAAGCGCCCTTTGTGTGAACATGCTGGCTGTGCAAA 180
OY 1034 ACATTGGCAATGAACAAAGTCTCACTAGCATGCTCTGTACATGATCTGCACAAGAG 1093
      |||||||
Db 181 ACATTGGCAATGAACAAAGTCTCACTAGCATGCTCTGTACATGATCTGCACAAGAG 240
OY 1094 AAATGAACTGATGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCATCTCACT 1153
      |||||||
Db 241 AAATGAACTGATGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCATCTCACT 300
OY 1154 GGATATATATCCCTCCCAAAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1213
      |||||||
Db 301 GGATATATATCCCTCCCAAAAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
OY 1214 TCACCCCACTGTGTGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCACT 1273
      |||||||
Db 361 TCACCCCACTGTGTGGAAGACAAAGTCTCGACACTGACACTACCTTGGCTCTCACT 420
```

RESULT 4

```
US-10-033-528-689
: Sequence 689, Application us/10033528
: Patent No. US20020131971A1
: GENERAL INFORMATION:
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jianshun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.547C1
: CURRENT APPLICATION NUMBER: US/10/033.528
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 689
: LENGTH: 439
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-033-528-689
```

```
Query Match
```

```
32.7%: Score 416.8; DB 12; Length 439;
```


Best Local Similarity 99.5%; Pred. No. 2.3e-102;
Matches 418; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 854 AAGCGAAGATTACCTTAAGCAACATGAAAACTCAGCCCGAGAAAGGATGTATGT 913
      |||
      1 AAGCGAAGATTACCTTAAGCAACATGAAAACTCAGCCCGAGAAAGGATGTATGT 60
OY 914 GCCTTCAGAGAAAGGCTGTGGAAGACCTATACAGTGTGTATCTCAAGCCAT 973
      |||
      61 CCCTTCAGAGAAAGGCTGTGGAAGACCTATACAGTGTGTATCTCAAGCCAT 120
OY 974 ATCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGTGACATCTGCTGGCAAA 1033
      |||
      121 ATCTCTCTCTCCATGAGAAAGCCGCTTTGTGTGTGACATCTGCTGGCAAA 180
OY 1034 ACATTTGCATGAACAAGTCTCAGTACGATGCTGTGTATCATGATCTGACAAG 1093
      |||
      181 ACATTTGCATGAACAAGTCTCAGTACGATGCTGTGTATCATGATCTGACAAG 240
OY 1094 AATATGAAGCTCAAGATCAAAAAATCTGTGAAAAACGAGTTGGCTCTCATCTCACT 1153
      |||
      241 AATATGAAGCTCAAGATCAAAAAATCTGTGAAAAACGAGTTGGCTCTCATCTCACT 300
OY 1154 GGATATATCTCTCCCAAGAAAGCAAGGAGCTTATCTTTGTGTCAAAAGGAGAG 1213
      |||
      301 GGATATATCTCTCCCAAGAAAGCAAGGAGCTTATCTTTGTGTCAAAAGGAGAG 360
OY 1214 TCACCCAACTGTGTGGAAGACAGATCTCTCGACATGTCAGTACCTTGGCTAA 1273
      |||
      361 TCACCCAACTGTGTGGAAGACAGATCTCTCGACATGTCAGTACCTTGGCTAA 420
DB
```

RESULT 5

US-09-796-692-4064

; Sequence 4064, Application US/09796692
; Publication No. US20020196362A1

; GENERAL INFORMATION:

; APPLICANT: Galger, Alexander

; APPLICANT: Mannion, Jane

; APPLICANT: Algate, Paul A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

; FILE REFERENCE: 2077.001200

; CURRENT APPLICATION NUMBER: US/09/796,692

; PRIORITY FILING DATE: 2001-03-01

; PRIORITY APPLICATION NUMBER: 60/186,126

; PRIORITY FILING DATE: 2000-03-01

; PRIORITY APPLICATION NUMBER: 60/190,479

; PRIORITY FILING DATE: 2000-03-17

; PRIORITY APPLICATION NUMBER: 60/200,545

; PRIORITY FILING DATE: 2000-04-27

; PRIORITY APPLICATION NUMBER: 60/200,303

; PRIORITY FILING DATE: 2000-04-28

; PRIORITY APPLICATION NUMBER: 60/200,779

; PRIORITY FILING DATE: 2000-04-28

; PRIORITY APPLICATION NUMBER: 60/200,999

; PRIORITY FILING DATE: 2000-05-01

; PRIORITY APPLICATION NUMBER: 60/202,084

; PRIORITY FILING DATE: 2000-05-04

; PRIORITY APPLICATION NUMBER: 60/206,201

; PRIORITY FILING DATE: 2000-05-22

; PRIORITY APPLICATION NUMBER: 60/218,950

; PRIORITY FILING DATE: 2000-07-14

; PRIORITY APPLICATION NUMBER: 60/222,903

; PRIORITY FILING DATE: 2000-08-03

; PRIORITY APPLICATION NUMBER: 60/223,416

; PRIORITY FILING DATE: 2000-08-04

; PRIORITY APPLICATION NUMBER: 60/223,378

; PRIORITY FILING DATE: 2000-08-07

; NUMBER OF SEQ ID NOS: 9597

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4064

; LENGTH: 234

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4064

Query Match 18.4%; Score 234; DB 9; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.9e-53;

Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 620 CTGAATTCATCACTGCGCAGCATACCAATGAACCTCTATTCAGTGTACCGAGAGGA 679
      |||
      1 CTGAATTCATCACTGCGCAGCATACCAATGAACCTCTATTCAGTGTACCGAGAGGA 60
OY 680 TGTGGGAACACTTTGCATCACCCAGCAGCTGAAGAGCATGCCAAGGCCAGAGGCG 739
      |||
      61 TGTGGGAACACTTTGCATCACCCAGCAGCTGAAGAGCATGCCAAGGCCAGAGGCG 120
OY 740 TATGATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATGAGCGAATCTTGAAACAT 799
      |||
      121 TATGATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATGAGCGAATCTTGAAACAT 180
OY 800 GTGAGGAACCCATTAAGAGAAATACTATGTGAATGATGCCGAAACATTT 853
      |||
      181 GTGAGGAACCCATTAAGAGAAATACTATGTGAATGATGCCGAAACATTT 234
DB
```

RESULT 6

US-09-920-300A-610

; Sequence 610, Application US/09920300A
; Patent No. US20020136728A1

; GENERAL INFORMATION:

; APPLICANT: King, Gordon E.

; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Xu, Jiaogchun

; TITLE OF INVENTION: SECRET, HEATHER

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.547

; CURRENT APPLICATION NUMBER: US/09/920,300A

; CURRENT FILING DATE: 2001-07-31

; NUMBER OF SEQ ID NOS: 1789

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 610

; LENGTH: 234

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-920-300A-610

Query Match 18.4%; Score 234; DB 10; Length 234;
Best Local Similarity 100.0%; Pred. No. 1.9e-53;

Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 620 CTGAATTCATCACTGCGCAGCATACCAATGAACCTCTATTCAGTGTACCGAGAGGA 679
      |||
      1 CTGAATTCATCACTGCGCAGCATACCAATGAACCTCTATTCAGTGTACCGAGAGGA 60
OY 680 TGTGGGAACACTTTGCATCACCCAGCAGCTGAAGAGCATGCCAAGGCCAGAGGCG 739
      |||
      61 TGTGGGAACACTTTGCATCACCCAGCAGCTGAAGAGCATGCCAAGGCCAGAGGCG 120
OY 740 TATGATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATGAGCGAATCTTGAAACAT 799
      |||
      121 TATGATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATGAGCGAATCTTGAAACAT 180
OY 800 GTGAGGAACCCATTAAGAGAAATCTATGTGAATGATGCCGAAACATTT 853
      |||
      181 GTGAGGAACCCATTAAGAGAAATCTATGTGAATGATGCCGAAACATTT 234
DB
```

RESULT 7

US-10-033-528-610

; Sequence 610, Application US/10033528
; Patent No. US20020131971A1

; GENERAL INFORMATION:


```
OTHER INFORMATION: spacer
NAME/KEY: misc.feature
LOCATION: (417)..(689)
OTHER INFORMATION: Three fingers of zinc fingers protein zlf268
NAME/KEY: misc.feature
LOCATION: (701)..(722)
OTHER INFORMATION: Nuclear Localization Signal
NAME/KEY: misc.feature
LOCATION: (957)..(986)
OTHER INFORMATION: c-myc tag
US-09-732-348-4

Query Match      12.0%  Score 152.4; DB 10; Length 995;
Best Local Similarity 57.6%  Pred. No. 3.7e-31;
Matches 273; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 268 GCGCGCGCGCGCGCTTCCAGAGGTTTCATCGCTCTCCCTCCGAGTGCAGGCCAATTA 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 26 GCGCGTGGCGGGTGTGTATGACCGGTACATCTGCTTTCGCGAAGTGGCGGCTGCTTA 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 CAGCAAAAGCTGGAACTTGAACGGGCACTGTGCAGCAACAGCAGCGGGAGAGACATTTGT 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 86 TTAACAAGAACTGGAAAGCGAGCGGCATCTGTCAAAACACAGAGAGAGAAACCATTTCC 145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 388 TTGTGACTATGAAGGCTGTGGCAAGGCTTCATCAGGAGCTACCATCTGAGCGGCACAT 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 146 ATGTAGAAGAGAGATGTGAGAAAGCTTTACCTGCTTCATCATTAAACCCCACTC 205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 448 TCTGACTCACACAGAGAAAGCCGTTTGTGACGCCACTGCTGTGATCAAAAAAT 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 ACTCACTATCTGGCGGAGAAACCTTCACATGTGACTGCGAGATGATGACTTGATTT 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 508 CAACCAAAATCAAACTTTGAGAAACATTTTGAACGAAACATGAACATCAACAAAAACA 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 266 TACTACAAAGGCAACATGAGAAAGCACTTTACAGATTCATACATACATATCTGGCT 325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 568 ATATATATGCACTTTTGAAGCTCTTAAGAAAGCTTTAAGAAACATCAGCAGCTGAAT 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 326 CTATGTGGCCATTTTGAAGCTGTGGCAAGCATTCAGAAACATCAATTTAAAGT 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 628 CCATCAGTGCAGCATATACCATTAACCTTATTCAAGTGTACCCAGAGAGATGTGGGA 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 386 TCATCAGTTCAAGTCAACACAGACGCTGCGTATGCTTGGCCCTGTGATCTCCTGGATCG 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 688 ACACGTTTGCATCACCACAGCAAGCTGAAAGCATGCCAAGGCCACGAGGCTTA 741
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 CCGCTTTTCTGCTGTGATGAGCTTACCGCGCATATCCGATCCACACAGGCCA 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-783-590-5443
Sequence 5443, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haselkline, William A.
APPLICANT: Li, Heodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
PRIOR FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 5443
LENGTH: 449
TYPE: DNA
ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: misc.feature
LOCATION: (8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (53)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (66)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (83)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (181)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (245)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (312)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (359)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (364)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (366)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (397)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (407)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc.feature
LOCATION: (448)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5443

Query Match      11.1%  Score 141.6; DB 10; Length 449;
Best Local Similarity 79.9%  Pred. No. 1.9e-28;
Matches 298; Conservative 0; Mismatches 58; Indels 17; Gaps 11;

QY 478 TTGTCCAGCCACTGCG-TGTGATCAAAATTTCAACACAAATCAAACTGAGAAACATT 536
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 63 TTGNCACAGCAATGGCTTGATATCAAAAATTCACACAAATCAAACTGAGAAACATT 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 537 TTG-TAACCAACATGAAATCAACAAACAAATATATATG-CAGTTTGAAGACTGTA 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 TTGAAGCGCAAAAGATGAAATCAACAAACAAATATATATGCGCACTTTGAAGACTGTA 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 595 -CAAGACCTTTAAGAAA-----CATCAGCAGCTGAAATTCATCATGCGCACATACCAA 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 GGAAGACCTTTAAGGAAACATCCAGAGCTGGAAATTCATCAGTGGCCAGCATACCAA 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 649 TG-AAACCTCTATT--CAAGTACCCA-GGAAGATGTGGGAAACACTTTG--CATCACCC 703
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 TGNAACTCTATTCCAAATGTACCCAGGGAAGATGTGGGAAACACTTTGGCAATCACCC 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 704 ACAGAGCTGAAA--CGACATGCAAAAGCCACAGAGGCTATGTATG-TCAAAAAGATGT 760
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 ACCAAGCTGMAAAGGACATGCCAAGGCCACAGAGGTTATGTATGTTCMAAAAGAGTGT 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 761 TCGTTTGGCAAAACATGACGAGCAACTCTGAAACATGTGAGAGAAACCATTAAGAG 820
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 TNCNTTGTGGCAAAACATGTGGAGGAGGACTTGNAACTGTGAGNAAACCTTTAAGGG 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 821 GAAATCTATGTG 833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 GATTATCTGTAG 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


THIS PAGE BLANK (USPTO)

Matches 1267; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY	1	ATGGCAGACACAGCGCGCGCCGACAGCGCGCGCTGCTGTGTGACGCGCGCGCTCCCGAAGT	60
Db	20	ATGGCAGACACAGCGCGCGCGACGCGCGCGCTGCTGTGTGACGCGCGCGCTCCCGAAGT	79
QY	61	GTGCGCGCGCTCGCCGAAAGTTAGCAGGAGACCTGGGCGCGCGCGCGCTTCCCGC	120
Db	80	GTGCGCGCGCTCGCGAAGTTTAGCAGGAGACCTGGGCGCGCGCGCTTCCCGC	138
QY	121	ACGTGTCTGGCAGCTGGCAGCGCGCTGGCCTTGGCTTGGAGGCGCGCGCGCTCGA	180
Db	139	ACGTGTCTGGCAGCTGGCAGCGCGCTGGCCTTGGCTTGGAGGCGCGCGCGCTCGA	198
QY	181	TGCGCGCGCGCTGTGCTGCCGAGTGCCTGTCTCTTACACATGCGCGACGCTTCATTGC	240
Db	199	TGCGCGCGCGCTGTGCTGCCGAGTGCCTGTCTCTTACACATGCGCGACGCTTCATTGC	258
QY	241	AGCGCGCGAGAGCTCACCTCCGACCCCGCGCGCGCGCGCGCTTCCGAGAGTATCTG	300
Db	259	AGCGCGCGAGAGCTCACCTCCGACCCCGCGCGCGCGCGCTTCCGAGAGTATCTG	318
QY	301	CTCTCTCCCTGACTGACGACCGCCCAATTACAGCAAAAGCTTGAAGCTTGTACGCGACCTGTG	360
Db	319	CTCTCTCCCTGACTGACGACCGCCCAATTACAGCAAAAGCTTGAAGCTTGTACGCGACCTGTG	378
QY	361	CACGACACAGGGGGAGAGACCATTTGTTGTGACTATGAAGGAGTGTGGCAGGCGCTGAT	420
Db	379	CACGACACAGGGGGAGAGACCATTTGTTGTGACTATGAAGGAGTGTGGCAGGCGCTGAT	438
QY	421	CAGGAGTACCATCTTGAGCGCGCCACATTTCTGACTCACAGAGAGAAAAAGCGCTTGTGG	480
Db	439	CAGGAGTACCATCTTGAGCGCGCCACATTTCTGACTCACAGAGAGAAAAAGCGCTTGTGG	498
QY	481	TGCAGCCACTGGCTGTGATCAAAATTCMAACCAAAATCAAACTTGAGAAACATTTTGA	540
Db	499	TGCAGCCAAATGGCTGTGATCAAAATTCMAACCAAAATCAAACTTGAGAAACATTTTGA	558
QY	541	ACGCAAAACATGAATAATGAACAAACAAATATATATGACGTTTTGGAGACCTGTAAAGAC	600
Db	559	ACGCAAAACATGAATAATGAACAAACAAATATATATGACGTTTTGGAGACCTGTAAAGAC	618
QY	601	CTTTAAGAAGCAGCAGACGTGAAATCCATCAGTGCAGCATACCAATGAACCTCTATT	660
Db	619	CTTTAAGAAGCAGCAGACGTGAAATCCATCAGTGCAGCATACCAATGAACCTCTATT	678
QY	661	CAAGGTATCCGGAAGAAGATGTGGGAAACATTTGCATACCCGACGAAGCTGAAGACACA	720
Db	679	CAAGGTATCCGGAAGAAGATGTGGGAAACATTTGCATACCCGACGAAGCTGAAGACACA	738
QY	721	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCTTTGTGCAAAAACATG	780
Db	739	TGCCAAGGCCACGAGGGCTATGTATGTCAAAAAGATGTTCTTTGTGCAAAAACATG	798
QY	781	GAGCGACCTGTGAAGATGTGAGAGAAACCCATTAAGAGAAATACTATATGGAAGTATG	840
Db	799	GAGCGACCTGTGAAGATGTGAGAGAAACCCATTAAGAGAAATACTATATGGAAGTATG	858
QY	841	CGGGAACATTTAAAGCAGAAATTAACCTTAAGCAACATGAGAAACTATGCCCGAGA	900
Db	859	CGGGAACATTTAAAGCAGAAATTAACCTTAAGCAACATGAGAAACTATGCCCGAGA	918
QY	901	AAGGAGTGTATGTGCGCTGTCCAGAGAAAGCGCTGGAAAGAACTATATCTATCTGTTTA	960
Db	919	AAGGAGTGTATGTGCGCTGTCCAGAGAAAGCGCTGTGGAAAGAACTATATCTATCTGTTTA	978
QY	961	TCTCCAAAGCATATCTCTCTCTTCATAGAGAAAGCGCGCTTTGTGTGAACATGC	1020
Db	979	TCTCCAAAGCATATCTCTCTCTTCATAGAGAAAGCGCGCTTTGTGTGAACATGC	1038
QY	1021	TGGGCTGTGGCAAAACATTTGCATGAACAAAGTGTCACTAGGATGCTGTTGTATATGA	1080
Db	1039	TGGGCTGTGGCAAAACATTTGCATGAACAAAGTGTCACTAGGATGCTGTTGTATATGA	1098

Qy	1081	TCTGACACAAGAAATGAAGTCACCAACTCAAAAATCTGTGAAATAAC-GGAGTTTGG	1139
Db	1099	TCTTGACACAGAATAATGAAGCTCAAAGTCAAAAAATCTGTAAATAACGGGAGTTTTGG	1158
Qy	1140	CCTCTCATCTCAGTGGATATATATCCCTCCCACAAAAGAACAGAGGCAGCTTATCTTTGGT	1199
Db	1159	CGCTCATCTCAGTGGATATAT-CTGCCAAAAGAAACAGAGGCAGCTTATCTTTGGT	1217
Qy	1200	GTCAAAACGAGAGTCAACCACACTGTGTGAAAGACAGATCCTTCGACAGTTGCAGTAC	1258
Db	1218	GTCAAAACGAGAGTCAACCACACTGTGTGAAAGACAGATCCTTCGACAGTTGCAGTAC	1277
Qy	1260	TTACCCCTGGCTAA	1273
Db	1278	TTACCCCTGGCTAA	1291

RESULT 2

US-08-523-376-1

Sequence 1, Application US/08523376

Patent No. 5808030

GENERAL INFORMATION:

APPLICANT: Satoshi TAKEDA

APPLICANT: YOSHIKAZU, SHIMADA

APPLICANT: Kouichi, OZAKI

APPLICANT: Sadahito, SIN

TITLE OF INVENTION: HTFIIIA GENE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS: ADDRESS: Subhrina

STREET: 2100 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: D.C.

; COUNTRY: United

ZIP: 20037-3202

COMPUTER READABLE
ADDITIONAL INFORMATION

MEDIUM TYPE: FLOPPY

OPERATING SYSTEM: PC-DOS

SOFTWARE: PatentIn Release

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/523,376

FILING DATE: _____

CLASSIFICATION: 536

TELECOMMUNICATIONS
TELEPHONE.

TELEPHONE: (202) 293-
TELEFAX: (202) 293-

TELEX: 6491103

; INFORMATION FOR SEQ ID NO:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1269 b

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: CDNA

MODECODE 11F; CDNA
US-08-523-376-1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query match	96.58
-------------	-------

Best Local Similarity

Matches 1264;

Query Match 96.50; Score 1228.6; DB 1; Length 1269;

Best Local Similarity 99.48; Pred. NO. 2.3e-311;

Matches 1264; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 1 ATGCGAGACAGCGGGCGCCGACGGCGCGCGTGGCTGTGTATCCGGCGCGCGCTCCGGAAAT 60
Db 1 ATGCGAGACAGCGGGCGCCGACGGCGCGCGTGGCTGTGTATCCGGCGCGCGCTCCGGAAAT 60
QY 61 GTGCGCGGCTCGCGCGAAGTTTCAGCAGGAGAGCGCTGTGGCGCGCGCGCGTTCGCCGAGC 120
Db 61 GTGCGCGGCTCGCGCGAAGTTTCAGCAGGAGAGCGCGTGTGGCGCGCGCGCGTTCGCCGAGC 119
QY 121 ACGTGTCTCGACACGTGGCAGCGCGCCTGGCGCTTGGAGGCGCGCGCGCCCTGGGA 180


```

Db 120 AGGTGCTGGGACGTGGACAGCGCCCTGGGCTTGAGGGCGCGCCCTGGGA 179
Qy 181 TCCGCCGCGCTGTGTGCCGAGGTGGTGTCTCTTGACATGCCAGCGCTTCATG 240
Db 180 TCCGCCGCGCGTGTGGCGAGGTGGTGTCTCTTGACATGCCAGCGCTTCATG 239
Qy 241 AGCGGCGGAGCTAGCTCGACCGCGCGCGCGCTTGAGAGGCTTCATG 300
Db 240 AGCGGCGGAGCTAGCTCGACCGCGCGCGCGCTTGAGAGGCTTCATG 299
Qy 301 CTCTCTCCCTGCTGACCGCGCAATTCAGAAAGCTTGAGCGCGACCTGTG 360
Db 300 CTCTCTCCCTGCTGACCGCGCAATTCAGAAAGCTTGAGCGCGACCTGTG 359
Qy 361 CAAGCAGCGGGGAGAGACCATTTGTTGATGATGAGGGGTGGCAAGGCTTCAT 420
Db 360 CAAGCAGCGGGGAGAGACCATTTGTTGATGATGAGGGGTGGCAAGGCTTCAT 419
Qy 421 CAGGGACTACATCTGAGCCGCGACATTTGACTCACAGAGAAAGCCGTTGTTG 480
Db 420 CAGGGACTACATCTGAGCCGCGACATTTGACTCACAGAGAAAGCCGTTGTTG 479
Qy 481 TGCAGCACTGCTGTGATCAAAAATTCACACAAATCAACTGAGAAACATTTGA 540
Db 480 TGCAGCACTGCTGTGATCAAAAATTCACACAAATCAACTGAGAAACATTTGA 539
Qy 541 ACCGAAACATGAAATCAACAAAACAAATATATGCAAGTTTGAAGACTGTAGAAG 600
Db 540 ACCGAAACATGAAATCAACAAAACAAATATATGCAAGTTTGAAGACTGTAGAAG 599
Qy 601 CTTTAAGAAACATCAGCAGCTGAAATTCATCAGTCAGCATACATGAACTGAT 660
Db 600 CTTTAAGAAACATCAGCAGCTGAAATTCATCAGTCAGCATACATGAACTGAT 659
Qy 661 CAAGTATACCCAGAAAGATGTGGAAACATTTGCATCACAGCAAGCTGAAGCACA 720
Db 660 CAAGTATACCCAGAAAGATGTGGAAACATTTGCATCACAGCAAGCTGAAGCACA 719
Qy 721 TGGCAAGGCGCAGAGGCTATGTATGTCAAAAAGATGTTCTTTGGCAAAAACATG 780
Db 720 TGGCAAGGCGCAGAGGCTATGTATGTCAAAAAGATGTTCTTTGGCAAAAACATG 779
Qy 781 GAGGAACTTTGAAACATGTGAGAAAGCAATTAAGAGAAATTCATGTGAACTATG 840
Db 780 GAGGAACTTTGAAACATGTGAGAAAGCAATTAAGAGAAATTCATGTGAACTATG 839
Qy 841 CCGGAAACATTTAAACCAAGATTAACCTTAAGCAACATGAAATCTATGCCGACA 900
Db 840 CCGGAAACATTTAAACCAAGATTAACCTTAAGCAACATGAAATCTATGCCGACA 899
Qy 901 AAGGATGATATGCTGCTGCAAGAGAGCTGTGAAAGCACTATTAAGTCTGTTTAA 960
Db 900 AAGGATGATATGCTGCTGCAAGAGAGCTGTGAAAGCACTATTAAGTCTGTTTAA 959
Qy 961 TCTCCAAACCAATATCTCTCTTCATGAGAAAGCGCGCTTTGTGTGTAACATGC 1020
Db 960 TCTCCAAACCAATATCTCTCTTCATGAGAAAGCGCGCTTTGTGTGTAACATGC 1019
Qy 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAGTCTCACTAGGATGCTGTTACATCA 1080
Db 1020 TGGCTGTGGCAAAACATTTGCAATGAACAAGTCTCACTAGGATGCTGTTACATCA 1079
Qy 1081 TCCGTGACAGAGAAATGAAGCTCAAAAGTCAAGTCAAGGCAAGGCTTATCTTTGT 1139
Db 1080 TCCGTGACAGAGAAATGAAGCTCAAAAGTCAAGTCAAGGCAAGGCTTATCTTTGT 1139
Qy 1140 CCGCTGATCTCAGTGGATATATCCCTCCCAAGAAACAAGGCAAGGCTTATCTTTGT 1199
Db 1140 CCGCTGATCTCAGTGGATATATCCCTCCCAAGAAACAAGGCAAGGCTTATCTTTGT 1198
Qy 1200 GTCAAAACGAGAGATCACCAACTGTGTGAAGCAAGATGCTCTGACAGTTGAGTAC 1259
Db 1199 GTCAAAACGAGAGATCACCAACTGTGTGAAGCAAGATGCTCTGACAGTTGAGTAC 1258

```

```

Qy 1260 TTACCTTGGC 1270
Db 1259 TTACCTTGGC 1269

RESULT 3
US-09-492-985-11
; Sequence 11, Application US/09492985
; Patent No. 6376240
; GENERAL INFORMATION:
; APPLICANT: Song, An M.
; APPLICANT: Chen, Ya-Fen
; APPLICANT: Krensky, Alan M.
; TITLE OF INVENTION: REPEAT-1: A Transcription Factor That
; FILE REFERENCE: SUN-113P
; CURRENT APPLICATION NUMBER: US/09/492,985
; EARLIER FILING DATE: 2000-01-27
; EARLIER FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO. 11
; LENGTH: 1471
; TYPE: DNA
; ORGANISM: mouse
US-09-492-985-11

Query Match 5.0%; Score 63.2; DB 4; Length 1471;
Best local similarity 48.9%; Pred. No. 4.2e-07;
Matches 170; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Qy 127 CTGCGACCTGGCAGCGGCGCTGGGCTTGAGAGGCGCGCGCTGATCCGC 185
Db 715 CCGGGAACAAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 774
Qy 187 GCGCGGTGCTGCGCGAGGTGGTGTCTCTTACCATGCGGACCGCTTCATTCAGCGC 246
Db 775 GGCATTGGAGCAGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 834
Qy 247 CGAGACTACGCTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 835 TCGGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 894
Qy 307 CCTGACTCAGCGCGCAATTAAGCAAGCTTGAAGCTTGAAGCGGACCTGTGCAAGCA 366
Db 895 CCGGCGCTCGGAAAGTTTACGGAATCTTCGACCTCAAGCGGACCTGGAATCTCA 954
Qy 367 CAGCGGGGAGACCATTTGTTGTGACTATGAAGGTTGTGCAAGGCTTCATCAGGCA 426
Db 955 CACAGGTGAGAGCGCTTTCGCTGCGAGCTGCGAGAGTGCACAAAGATTTCCACGCTC 1014
Qy 427 CTACCATCTGACCGCGCACTTCACTACACAGAGAAAGCGCTT 474
Db 1015 GAGCAGACTGCGCAGCGCACTATGCGACGCAAGCGGGAAGAGATT 1062

RESULT 4
US-09-492-985-1
; Sequence 1, Application US/09492985
; Patent No. 6376240
; GENERAL INFORMATION:
; APPLICANT: Song, An M.
; APPLICANT: Chen, Ya-Fen
; APPLICANT: Krensky, Alan M.
; TITLE OF INVENTION: REPEAT-1: A Transcription Factor That
; FILE REFERENCE: SUN-113P
; CURRENT APPLICATION NUMBER: US/09/492,985
; EARLIER FILING DATE: 2000-01-27
; EARLIER FILING DATE: 1999-01-27

```

NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1430
TYPE: DNA
ORGANISM: homo sapien
FEATURE:
OTHER INFORMATION: synthetic oligonucleotide probe
US-09-492-985-1

Query Match 4.8%; Score 61.4; DB 4; Length 1430;
Best Local Similarity 48.9%; Pred. No. 1.2e-06;
Matches 193; Conservative 0; Mismatches 201; Indels 1; Gaps 1;

81 TTCACAGAGGAGCCGTGGCCGCGCGGCGGCTTCCGCGAGCTGTCTCTGCGACGTGGCA 140
Db TCCCGCGCTGCG 723
141 GCGCCGCTGCG 199
Db GCG 783
200 GACGTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 259
Db GACCG 843
260 CCGACCG 319
Db ACTCG 903
320 GCCAATTACAGCAAAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 379
Db AAGATTACAGCAAAAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 963
380 CCATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 439
Db CCCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1023
440 CGCCACATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 474
Db CGCAGCTACCGCGACACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1058

RESULT 5
US-09-362-123A-3
Sequence 3, Application US/09362123A
Patent No. 6451558
GENERAL INFORMATION:
APPLICANT: Cooke, Michael Paul
APPLICANT: Holmes, Claire
TITLE OF INVENTION: No. 6451558el Genes In the Control of Hematopoiesis
FILE REFERENCE: 4-306294/SYS
CURRENT APPLICATION NUMBER: US/09/362,123A
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: 09/128,310
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 3
LENGTH: 2992
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (82)..(1710)
US-09-362-123A-3

Query Match 4.6%; Score 58.8; DB 4; Length 2992;
Best Local Similarity 54.6%; Pred. No. 8e-06;
Matches 142; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

QY 238 TGCAGCCGCGGAGAGTACTGCTCCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db TCCCTTCAGCCGCGGCTCTGCTTTAAGCAGCAGCGCGCTGCACACAGCGGAGAAAGCC 1322
QY 298 CTGCTCTCTCCCTGACTGACGCGCAATACAGCAAAAGCTGAGAGCTTGAAGCGCACT 357
Db TCCGCGCTGCGCAGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1382
QY 358 GTGCAGCAGCAGCGGAGAGACCATTTGTTGACTATGAAGGTGTGGCAAGCCTT 417
Db GTCTCTGACAGAGGCG 1436
QY 418 CATTGAGGAGTACATCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477
Db CCGCAAGGCG 1496
QY 478 TTGTGACGCACTGGCTGTG 497
Db TGTAGCGAGTGTGCGCGG 1516

RESULT 6
US-09-234-332-5
Sequence 5, Application US/09234332A
Patent No. 6087168
GENERAL INFORMATION:
APPLICANT: Cedars-Sinai Medical Center
APPLICANT: Michel F. Levesque, M.D.
TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO
FILE REFERENCE: P07 41494
CURRENT APPLICATION NUMBER: US/09/234,332A
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 3138
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: gene
LOCATION: (0)...(0)
OTHER INFORMATION: Zic 1 Protein gene; Genbank Accession D76435
US-09-234-332-5

Query Match 4.3%; Score 54.8; DB 3; Length 3138;
Best Local Similarity 55.2%; Pred. No. 9.1e-05;
Matches 107; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 341 AAGCTTGACCGCGCACTCTGCAACACACCGGGGAGAGACATTTTGTGACTATGAA 400
Db TCCCTTCAGCCGCGGCTCTGCTTTAAGCAGCAGCGCGCTGCACACAGCGGAGAAAGCC 1701
QY 401 GCGTGTGCAAGCGCTTATCAGGAGTACCATCTGAGCGCGCACTTGTGACTACACA 460
Db TCCCTTCAGCCGCGGCTCTGCTTTAAGCAGCAGCGCGCTGCACACAGCGGAGAAAGCC 1761
QY 461 GGAGAAAGCGCTTGTGTGACGACCATGCTGTGATCAAAATTCACAAATCA 520
Db GGAGAAAGCGCTTGTGTGACGACCATGCTGTGATCAAAATTCACAAATCA 520
QY 521 AACTGAAGAAACA 534
Db TCCCTTCAGCCGCGGCTCTGCTTTAAGCAGCAGCGCGCTGCACACAGCGGAGAAAGCC 1821
QY 1822 GACCGCAAGAAACA 1835

RESULT 7
US-08-570-227A-1
Sequence 1, Application US/08570227A
Patent No. 5981217
GENERAL INFORMATION:
APPLICANT: Subramaniam, M.

TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpc-fls
US-08-232-463-14

Query Match 4.1%: Score 52.8; DB 1; Length 7218;
Best Local Similarity 4.5%: Pred. No. 0.00043;
Matches 18; Conservative 210; Mismatches 160; Indels 0; Gaps 0;

511 CACAAATCAAACTTGAAGAACTTTGAACGCAAAACATGAAATCAACAAACATA 570
1466 CAAGTACTTAAAGATGAAGATTGTACRRRRRRRRRRRRRRRRRRRRRRRR 1407
571 TATATGCACTTTGAACCTCTAAGAACCTTTAAGAAACATCAACCTGAATAATCA 630
1406 RRR 1347
631 TCAGTCCAGCATACCACTCTATTCAAGTACCCAGCAGAGCTGGGAAACA 690
1346 RRR 1287
691 CTTTCATCACCAGCAAGCTGAAGACATGCCAAGCCAGGCGCTATGTATGCA 750
1286 RRR 1227
751 AAAGATGTCCTTGGCAAAACATGACGAGCACTTCTGAACATGTAGAGAAAC 810
1226 RRR 1167
811 CCATTAAGAGCAATACTATGTGAAGTATCCCGAAACATTTAAGCAAGATTACCT 870
1166 RRR 1107
871 TAAGCAACATGAACACTCATGCCCGAGAAAGGA 906
1106 RRR 1071

RESULT 10
US-09-657-042A-3
Sequence 3, Application US/09657042A
Patent No. 6329203
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Jacqueline Wyatt
TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-1 EXPRESSION
FILE REFERENCE: RTS-0148
CURRENT APPLICATION NUMBER: US/09/657,042A
CURRENT FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 88
SEQ ID NO 3
LENGTH: 3600
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (79) ... (3399)
US-09-657-042A-3

Query Match 4.1%: Score 52.6; DB 4; Length 3600;
Best Local Similarity 51.2%: Pred. No. 0.00036;
Matches 149; Conservative 0; Mismatches 139; Indels 3; Gaps 1;

341 AAGCTTGACGCGCACTGTGCAACGACGCGGGAGAGACATTTGTGACTATGAA 400
1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
937 ATGCTGTGTTACATGCGGACACACTGCGGAGAACACACAAAGTCGACGTTGAA 996

401 GGCTGCGCAAGCCCTTCATCAGGACTACCATCTGACGCCACATTTGTGACTACACA 460
997 GGGTCGCGCAAGCTATCTACTACAGCTCGAAGACCTGMAAGCAGCCTGGTCACACAG 1056
461 GGAGAAAGCCGTTGTTGTGACAGCAGCTGCTGTGATCAAAATTCACACAAATCA 520
1057 GGTGAAGAGCCATACATGTGTGACAGCAGAGGCTGACAGTAACCTTCAGCAATGCCAGT 1116
521 AACTTGAAAGAACTTTGAAGCGCAACATGAAATTCACAAACATATATATGACGT 580
1117 GACCGAGCGCAACGACAGCAAGATCGAGCCATTCCAAAT--GAGAACGCTATGTATGTAAG 1173
581 TTTGAAGACTGTAAAGAACCTTTAAGAAACATCAGCAGCTGCAAAATCCAT 631
1174 CTCCTCGGTGACCAACGCTATACAGATCTCAGCTCGGTGGAAACAT 1224

RESULT 11
US-09-907-843-3
Sequence 3, Application US/09907843
Patent No. 6440739
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Susan M. Freiler
TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-2 EXPRESSION
FILE REFERENCE: RTS-0279
CURRENT APPLICATION NUMBER: US/09/907,843
CURRENT FILING DATE: 2001-07-17
NUMBER OF SEQ ID NOS: 87
SEQ ID NO 3
LENGTH: 4960
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (572) ... (4348)
US-09-907-843-3

Query Match 4.0%: Score 51; DB 4; Length 4960;
Best Local Similarity 52.5%: Pred. No. 0.0011;
Matches 136; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

341 AAGCTTGACCGCCACCTGTGCAACGACGCGGAGAGACCATTTGTTGTACTATGAA 400
1052 ATGCTGTGTTGTCACATCTCCGCGACACGCGGAGAGCCGACAAAGTCGACGTTGAG 1111
401 GGCTGCGCAAGCCCTTCATCAGGACTACCATCTGAGCGCGCACATTTGTGACTACACA 460
1112 GGCTGCTGGAAGGCTACTCTCCGCTGAGAACCTGGAAGACACCTGGGTCCACAC 1171
461 GGAGAAAGCCGTTGTTGTGACAGCAGCTGCTGTGATCAAAATTCACAAATCA 520
1172 GGGGAGAAAGCATATGTGTGACACGAGGCTGCAACAAACCTTTCGAAACGCTCG 1231
521 AACTTGAAAGAACTTTGAAGCGCAACATGAAATTCACAAACATATATATGACGT 580
1232 GACCGCGCAACGACAGAAATGCGACCACTCCA---ACGAGAAACCTTACATGTGCAAG 1288
581 TTTGAAGACTGTAAAGAGA 599
1289 ATCCAGGCTGACCAAGA 1307

RESULT 12
US-08-946-241B-1
Sequence 1, Application US/08946241B
Patent No. 5928941
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: MCA/Nulty, Megan M.
TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:

```

ADDRESS: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,241B
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creason, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 434...1843
US-08-946-241B-1

Query Match
Best Local Similarity 55.0%; Score 50.4; DB 2; Length 1889;
Matches 99: Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 295 CATCTGCTCCTTCCCTGATCGACGCCCAATTCAGCAAGAGCTGGAAGCTTGAGCGCA 354
DB 1594 CACTTGATGATTACGGGGCTGCGCAAAACCTACACAAAGATTCCCATCTCAAGGCACA 1653
QY 355 CCTGTGCAAGCAGACGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGTGCAAGGC 414
DB 1654 CTGTGGAACCCACAGAGGTGAGAAACCTTACCACTGTGACTGGAGGCTGTGATGGA 1713
QY 415 CTTCATCAGGAGACTACATCTGAGCCGCCACATTTGACTCAGACAGGAAAAAGCCGTT 474
DB 1714 ATTGCGCCGCTCAGATGAGTGAACGACGACACTACCGTAACACACGGGCGCCCGCTT 1773

RESULT 13
US-08-946-241B-8
Sequence 8, Application US/08946241B
Patent No. 5928941
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: MCA/Nulty, Megan M.
TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
```

```

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,241B
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creason, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 407...1843
US-08-946-241B-8

Query Match
Best Local Similarity 55.0%; Score 50.4; DB 2; Length 1889;
Matches 99: Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 295 CATCTGCTCCTTCCCTGATCGACGCCCAATTCAGCAAGAGCTGGAAGCTTGAGCGCA 354
DB 1594 CACTTGATGATTACGGGGCTGCGCAAAACCTACACAAAGATTCCCATCTCAAGGCACA 1653
QY 355 CCTGTGCAAGCAGACGGGGAGAGACCATTTGTTGTGACTATGAGAGGTGTGCAAGGC 414
DB 1654 CTGTGGAACCCACAGAGGTGAGAAACCTTACCACTGTGACTGGAGGCTGTGATGGA 1713
QY 415 CTTCATCAGGAGACTACATCTGAGCCGCCACATTTGACTCAGACAGGAAAAAGCCGTT 474
DB 1714 ATTGCGCCGCTCAGATGAGTGAACGACGACACTACCGTAACACACGGGCGCCCGCTT 1773

RESULT 14
US-09-309-053-1
Sequence 1, Application US/09309053
Patent No. 6077933
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: MCA/Nulty, Megan M.
TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/309,053
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/946,241
```

FILED DATE: 07-OCT-1997
APPLICATION NUMBER: 60/030,035
FILING DATE: 05-NOV-1996
APPLICATION NUMBER: 60/027,521
FILING DATE: 07-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Creason, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 434...1843
US-09-309-053-1

Query Match 4.0%; Score 50.4; DB 3; Length 1889;
Best Local Similarity 55.0%; Pred. No. 0.001; Indels 0; Gaps 0;

Matches 99; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 295 CATCTGCTCCTTCCTGACTGACGCGCAATTCACAGCAAGCCTGGAAGCTTGACGGCA 354
DB 1594 CACTTGATTTACGGGGCTCGGCAAAACCTACACAAAGAGTTCATCTCAAGGCACA 1653
QY 355 CCTGTGCAAGACACAGGGGAGAGACCATTTGTTGTGACTATGAAGGTGTGGCAAGC 414
DB 1654 CCTGGAAACCCACACAGGTGAGAAACCTTACCACTGTGACTGGAGCGGTGGATGAA 1713
QY 415 CTTGATCAGGAGCTACCATCTGACCGCCACATTCGACTCACACAGAGAAACCGCTT 474
DB 1714 ATTCGCCCGCTCAGATGACTGACGAGCAGCACTACCGTAAACACAGGGGACCGCCCTT 1773

RESULT 15
US-09-309-053-8

Sequence 8, Application US/09309053
Patent No. 6077933

GENERAL INFORMATION:

APPLICANT: Lee, Mu-En

APPLICANT: MCA/Nuilty, Megan M.

TITLE OF INVENTION: REPRESSOR KRUPPEL-LIKE FACTOR

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/309,053

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/946,241

FILING DATE: 07-OCT-1997

APPLICATION NUMBER: 60/030,035

FILING DATE: 05-NOV-1996

APPLICATION NUMBER: 60/027,521

FILING DATE: 07-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Creason, Gary L.
REGISTRATION NUMBER: 34,310
REFERENCE/DOCKET NUMBER: 05433/027001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1889 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 407...1843
US-09-309-053-8

Query Match 4.0%; Score 50.4; DB 3; Length 1889;
Best Local Similarity 55.0%; Pred. No. 0.001; Indels 0; Gaps 0;

Matches 99; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 295 CATCTGCTCCTTCCTGACTGACGCGCAATTCACAGCAAGCCTGGAAGCTTGACGGCA 354
DB 1594 CACTTGATTTACGGGGCTCGGCAAAACCTACACAAAGAGTTCATCTCAAGGCACA 1653
QY 355 CCTGTGCAAGACACAGGGGAGAGACCATTTGTTGTGACTATGAAGGTGTGGCAAGC 414
DB 1654 CCTGGAAACCCACACAGGTGAGAAACCTTACCACTGTGACTGGAGCGGTGGATGAA 1713
QY 415 CTTGATCAGGAGCTACCATCTGACCGCCACATTCGACTCACACAGAGAAACCGCTT 474
DB 1714 ATTCGCCCGCTCAGATGACTGACGAGCAGCACTACCGTAAACACAGGGGACCGCCCTT 1773

Search completed: February 10, 2003, 12:47:09

Job time: 111.743 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2003, 12:44:26 : Search time 3124.12 seconds
(without alignments)
10244.915 Million cell updates/sec

Title: US-09-831-426C-3
Perfect score: 1273
Sequence: 1 atgcgcagcagcgcgcgca.....cagctacttacctgtctaa 1273

Scoring table: IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending-Patents-NA_Main:*
1: /cgn2_6/prodata/1/pna/US06.COMB.seq:*
2: /cgn2_6/prodata/1/pna/US06.COMB.seq:*
3: /cgn2_6/prodata/1/pna/US07.COMB.seq:*
4: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
5: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
6: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
7: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
8: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
9: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
10: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
11: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
12: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
13: /cgn2_6/prodata/1/pna/US08.COMB.seq:*
14: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
15: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
16: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
17: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
18: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
19: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
20: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
21: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
22: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
23: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
24: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
25: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
26: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
27: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
28: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
29: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
30: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
31: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
32: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
33: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
34: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
35: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
36: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
37: /cgn2_6/prodata/1/pna/US09.COMB.seq:*
38: /cgn2_6/prodata/1/pna/US10.COMB.seq:*
39: /cgn2_6/prodata/1/pna/US10.COMB.seq:*
40: /cgn2_6/prodata/1/pna/US10.COMB.seq:*
41: /cgn2_6/prodata/1/pna/US10.COMB.seq:*
42: /cgn2_6/prodata/1/pna/US10.COMB.seq:*
43: /cgn2_6/prodata/1/pna/US10.COMB.seq:*

SUMMARIES					Result		Query		Description	
No.	Score	% Match	Length	ID	No.	Score	% Match	Length	ID	Description
1	1273	100.0	1273	32	US-09-831-426-1	44:	/cgn2_6/prodata/1/pna/US6000.COMB.seq.*		Sequence 1, Appl1	
2	1273	100.0	1273	32	US-09-831-426-3	45:	/cgn2_6/prodata/1/pna/US6001.COMB.seq.*		Sequence 3, Appl1	
3	1273	100.0	1273	32	US-09-831-426C-1	46:	/cgn2_6/prodata/1/pna/US6002.COMB.seq.*		Sequence 1, Appl1	
4	1273	100.0	1273	32	US-09-831-426C-1	47:	/cgn2_6/prodata/1/pna/US6003.COMB.seq.*		Sequence 3, Appl1	
5	1269.8	99.7	1597	25	US-09-652-126-8862	48:	/cgn2_6/prodata/1/pna/US6004.COMB.seq.*		Sequence 8862, Ap	
6	1269.8	99.7	1597	25	US-09-652-126-8862	49:	/cgn2_6/prodata/1/pna/US6005.COMB.seq.*		Sequence 16879, A	
7	1269.8	99.7	1597	25	US-09-652-814-10703	50:	/cgn2_6/prodata/1/pna/US6006.COMB.seq.*		Sequence 18058, A	
8	1269.8	99.7	1597	25	US-09-652-814-10703	51:	/cgn2_6/prodata/1/pna/US6007.COMB.seq.*		Sequence 4, Appl1	
9	1269.8	99.7	1597	25	US-09-716-473-1934	52:	/cgn2_6/prodata/1/pna/US6008.COMB.seq.*		Sequence 1668, Ap	
10	1269.8	99.7	1597	25	US-09-716-473-1934	53:	/cgn2_6/prodata/1/pna/US6009.COMB.seq.*		Sequence 1678, Ap	
11	1269.8	99.7	1597	25	US-09-721-589-6336	54:	/cgn2_6/prodata/1/pna/US6010.COMB.seq.*		Sequence 3560, Ap	
12	1246.2	97.9	2519	76	US-60-324-185-16829	55:	/cgn2_6/prodata/1/pna/US6011.COMB.seq.*		Sequence 3560, Ap	
13	1222.6	96.0	1527	32	US-60-172-373-18068	56:	/cgn2_6/prodata/1/pna/US6012.COMB.seq.*		Sequence 18056, A	
14	1211.4	95.2	1213	32	US-09-831-426-4	57:	/cgn2_6/prodata/1/pna/US6013.COMB.seq.*		Sequence 24482, A	
15	1211.4	95.2	1213	32	US-09-831-426-4	58:	/cgn2_6/prodata/1/pna/US6014.COMB.seq.*			
16	1206	94.7	1349	1	PCT-US00-26524B-1668	59:	/cgn2_6/prodata/1/pna/US6015.COMB.seq.*			
17	1206	94.7	1349	1	PCT-US00-26524B-1668	60:	/cgn2_6/prodata/1/pna/US6016.COMB.seq.*			
18	1187.6	93.3	1309	28	US-09-705-256A-3560	61:	/cgn2_6/prodata/1/pna/US6017.COMB.seq.*			
19	1187.6	93.3	1309	28	US-09-705-256A-3560	62:	/cgn2_6/prodata/1/pna/US6018.COMB.seq.*			
20	983.6	77.3	1220	61	US-60-172-373-18066	63:	/cgn2_6/prodata/1/pna/US6019.COMB.seq.*			
21	774.6	60.8	802	16	US-09-205-070-24482	64:	/cgn2_6/prodata/1/pna/US6020.COMB.seq.*			
						65:	/cgn2_6/prodata/1/pna/US6021.COMB.seq.*			
						66:	/cgn2_6/prodata/1/pna/US6022.COMB.seq.*			
						67:	/cgn2_6/prodata/1/pna/US6023.COMB.seq.*			
						68:	/cgn2_6/prodata/1/pna/US6024.COMB.seq.*			
						69:	/cgn2_6/prodata/1/pna/US6025.COMB.seq.*			
						70:	/cgn2_6/prodata/1/pna/US6026.COMB.seq.*			
						71:	/cgn2_6/prodata/1/pna/US6027.COMB.seq.*			
						72:	/cgn2_6/prodata/1/pna/US6028.COMB.seq.*			
						73:	/cgn2_6/prodata/1/pna/US6029.COMB.seq.*			
						74:	/cgn2_6/prodata/1/pna/US6030.COMB.seq.*			
						75:	/cgn2_6/prodata/1/pna/US6031.COMB.seq.*			
						76:	/cgn2_6/prodata/1/pna/US6032.COMB.seq.*			
						77:	/cgn2_6/prodata/1/pna/US6033.COMB.seq.*			
						78:	/cgn2_6/prodata/1/pna/US6034.COMB.seq.*			
						79:	/cgn2_6/prodata/1/pna/US6035.COMB.seq.*			
						80:	/cgn2_6/prodata/1/pna/US6036.COMB.seq.*			
						81:	/cgn2_6/prodata/1/pna/US6037.COMB.seq.*			
						82:	/cgn2_6/prodata/1/pna/US6038.COMB.seq.*			
						83:	/cgn2_6/prodata/1/pna/US6039.COMB.seq.*			
						84:	/cgn2_6/prodata/1/pna/US6040.COMB.seq.*			
						85:	/cgn2_6/prodata/1/pna/US6041.COMB.seq.*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```
22 774.6 60.8 802 17 US-09-340-623-24482 Sequence 24482, A
23 774.6 60.8 802 33 US-09-898-888A-24482 Sequence 24482, A
24 774.6 60.8 802 33 US-09-898-888A-24482 Sequence 24482, A
25 613 48.2 776 25 US-09-652-814-6814 Sequence 6814, Ap
26 555.4 43.6 1359 20 US-09-535-897-7605 Sequence 7605, Ap
27 546.4 42.9 1359 27 US-09-644-873-10092 Sequence 10092, A
28 546.4 42.9 1359 27 US-09-699-998-9945 Sequence 9545, Ap
29 541.2 42.5 1467 80 US-60-360-207-7415 Sequence 7415, Ap
30 495 38.9 690 1 PCT-US01-18569-803 Sequence 803, App
31 495 38.9 690 43 US-10-264-049-803 Sequence 803, App
32 470.8 37.0 969 25 US-09-644-873-7308 Sequence 7308, Ap
33 469.2 36.9 497 17 US-09-306-603-4878 Sequence 4878, Ap
34 469.2 36.9 497 19 US-09-522-251-4878 Sequence 4878, Ap
35 469.2 36.9 497 18 US-10-025-600-4878 Sequence 4878, Ap
36 453.4 35.6 618 20 US-09-535-897-7597 Sequence 7597, Ap
37 429.2 33.7 650 81 US-60-377-240-6706 Sequence 6706, Ap
38 418.4 32.9 439 34 US-09-920-300-326 Sequence 326, App
39 418.4 32.9 439 38 US-09-920-300A-326 Sequence 326, App
40 418.4 32.9 439 38 US-10-033-528-326 Sequence 326, App
41 418.4 32.9 439 39 US-10-099-928-326 Sequence 326, App
42 416.8 32.7 439 34 US-09-920-300-689 Sequence 689, App
43 416.8 32.7 439 34 US-09-920-300A-689 Sequence 689, App
44 416.8 32.7 439 38 US-10-033-528-689 Sequence 689, App
45 416.8 32.7 439 39 US-10-099-928-689 Sequence 689, App
```

ALIGNMENTS

```
RESULT 1
US-09-831-426-1
: Sequence 1, Application US/09831426
: GENERAL INFORMATION:
: APPLICANT: Hoechst Marion Roussel
: APPLICANT: Borden-Pallier, F.
: APPLICANT: Rochem, C.
: TITLE OF INVENTION: Human htf111A gene and coded htf111A protein
: FILE REFERENCE: 146.1364
: CURRENT APPLICATION NUMBER: US/09/831,426
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Vers. 2.0
: SEQ ID NO 1
: LENGTH: 1273
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (176)..(1270)
US-09-831-426-1
```

```
Query Match Best Local Similarity 100.0%; Score 1273; DB 32; Length 1273;
Matches 1273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGCGCAGCAGCGCGCGCGACGCGGCGGTGCTGTCACCGCGCGCTCCCGAAGT 60
Db 1 ATGCGCAGCAGCGCGCGCGCGACGCGGCGGTGCTGTCACCGCGCGCTCCCGAAGT 60
QY 61 GTGCGCGGCTCGCGGGAAGTTACAGAGAGAGCGCTGCGCGCGCGCGCTTCCGCG 120
Db 61 GTGCGCGGCTCGCGGGAAGTTACAGAGAGAGCGCTGCGCGCGCGCGCTTCCGCG 120
QY 121 ACCTGCTCTGCGACGTGCGACGCGCGCTGCGCTGCGCTTGAAGAGCGCGCGCTTGA 180
Db 121 ACCTGCTCTGCGACGTGCGACGCGCGCTGCGCTGCGCTTGAAGAGCGCGCGCTTGA 180
QY 181 TCCGCGCGCGCGGTGCGCGCGAGCGCTGCTGCTGACCATGCGCGAGCGCTTATTGTC 240
Db 181 TCCGCGCGCGCGGTGCGCGCGAGCGCTGCTGCTGACCATGCGCGAGCGCTTATTGTC 240
QY 241 ACCGCGGAGAGCTTACGCTCCGACCCCGCGCGCGCGCTTCCAGAGAGTTATCTG 300
Db 241 ACCGCGGAGAGCTTACGCTCCGACCCCGCGCGCGCGCTTCCAGAGAGTTATCTG 300
```

```
Db 241 ACCGCGGAGAGCTTACGCTCCGACCCCGCGCGCGCGCTTCCAGAGAGTTATCTG 300
QY 301 CTCCTTCCCTGACGTGCGACGCGCAATTAACAGCAAGCGCTGGAAGCTTGACGCGCCTGTG 360
Db 301 CTCCTTCCCTGACGTGCGACGCGCAATTAACAGCAAGCGCTGGAAGCTTGACGCGCCTGTG 360
QY 361 CAAGCAACGCGGAGAGACCATTTGTTGTGACTATGAAGGTGCGCAAGCGCTTAT 420
Db 361 CAAGCAACGCGGAGAGACCATTTGTTGTGACTATGAAGGTGCGCAAGCGCTTAT 420
QY 421 CAGGAGTACCATCTGACGCGCGCAATTTGACTCACACAGAGAAAGCCGTTGTTG 480
Db 421 CAGGAGTACCATCTGACGCGCGCAATTTGACTCACACAGAGAAAGCCGTTGTTG 480
QY 481 TCGACGCACTGCTGTGATCAAAAATTCACACAAATCAACTGGAAGAAACATTTTGA 540
Db 481 TCGACGCACTGCTGTGATCAAAAATTCACACAAATCAACTGGAAGAAACATTTTGA 540
QY 541 ACCGCAACATGAANAATCAACAAACAAATATATATGCAAGTTTGAAGACTGTAAGAAC 600
Db 541 ACCGCAACATGAANAATCAACAAACAAATATATATGCAAGTTTGAAGACTGTAAGAAC 600
QY 601 CTTTAAGAAACATCAGCAGCTGAATAATCATCATGTCAGTACATACATGAACCTTATT 660
Db 601 CTTTAAGAAACATCAGCAGCTGAATAATCATCATGTCAGTACATACATGAACCTTATT 660
QY 661 CAAGTGTACCCAGAGAGATGTGGGAAACACTTGCATCACCAGCAAGCTAAACGCA 720
Db 661 CAAGTGTACCCAGAGAGATGTGGGAAACACTTGCATCACCAGCAAGCTAAACGCA 720
QY 721 TCGCAAGCGCCAGAGGCGCTATGATGTCAAAAAGATGTCCTTTGTGGCAAAACATG 780
Db 721 TCGCAAGCGCCAGAGGCGCTATGATGTCAAAAAGATGTCCTTTGTGGCAAAACATG 780
QY 781 GACGCACTTCTGAAACATGTGACAGAAACCCATTAAGAGAAATCTATGGAAGTATG 840
Db 781 GACGCACTTCTGAAACATGTGACAGAAACCCATTAAGAGAAATCTATGGAAGTATG 840
QY 841 CCGGAAACATTTAAACCAAGATATACCTTAAGCAACATGAACATGATGCCCCAGA 900
Db 841 CCGGAAACATTTAAACCAAGATATACCTTAAGCAACATGAACATGATGCCCCAGA 900
QY 901 AAGGATGATGTGCTGCTGCTCCAGAGAGGCTGGAAGAACTATACATGCTGTTTAA 960
Db 901 AAGGATGATGTGCTGCTGCTCCAGAGAGGCTGGAAGAACTATACATGCTGTTTAA 960
QY 961 TCTCAAAAGCCATATCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1020
Db 961 TCTCAAAAGCCATATCTCTCTCCATGAGAGAAAGCGCCCTTTGTGTGAACATGC 1020
QY 1021 TGGCTGTGGCAAAACATTTGAAGAAACAAACATGCTACATGACATGCTGTCATGA 1080
Db 1021 TGGCTGTGGCAAAACATTTGAAGAAACAAACATGCTACATGACATGCTGTCATGA 1080
QY 1081 TCGTGAACAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGAAGAAAGAGTTTGGC 1140
Db 1081 TCGTGAACAAGAAATGAAGCTCAAAAGTCAAAAATCTGCTGAAGAAAGAGTTTGGC 1140
QY 1141 CTCTCATCTCACTGATATATCCCTCCCAAGAAAGAAAGGCAAGGCTTATCTTTGTG 1200
Db 1141 CTCTCATCTCACTGATATATCCCTCCCAAGAAAGAAAGGCAAGGCTTATCTTTGTG 1200
QY 1201 TCAAAAAGAGAGTACACCACTGTGTGAAGAGCAAGATGCTTCAGAGTTCAGTACT 1260
Db 1201 TCAAAAAGAGAGTACACCACTGTGTGAAGAGCAAGATGCTTCAGAGTTCAGTACT 1260
QY 1261 TACCTTGGCTTAA 1273
Db 1261 TACCTTGGCTTAA 1273
RESULT 2
US-09-831-426-3
```


Db 121 AGGTCTGCGCAGCTGGCAGCCGCTGGCCCTGGGCTTGGAGCGCCCGCCCTGGA 180
Qy 181 TCCGCGCGCGCTGGTGGCGGAGTGGTGGTCTTGGACATCCCGAGCGCTTCATTGGC 240
Db 181 TCCGCGCGCGCTGGTGGCGGAGTGGTGGTCTTGGACATCCCGAGCGCTTCATTGGC 240
Qy 241 ACCCGGCGAGAGCTCAGCTCCGAGCCCGCGCGCGCTTCCGAGAGGTTTCATCTG 300
Db 241 ACCCGGCGAGAGCTCAGCTCCGAGCCCGCGCGCGCTTCCGAGAGGTTTCATCTG 300
Qy 301 CTCCTTCCCTACTGTCAGCGCAATTAAGCAAAAGCCTTGAAGCTTGAACGCGCATGTG 360
Db 301 CTCCTTCCCTACTGTCAGCGCAATTAAGCAAAAGCCTTGAAGCTTGAACGCGCATGTG 360
Qy 361 CAGGACACGGGGGAGAGACCATTTGTTGTGATGAAGGTTGGCAGAGCCTTCAT 420
Db 361 CAGGACACGGGGGAGAGACCATTTGTTGTGATGAAGGTTGGCAGAGCCTTCAT 420
Qy 421 CAGGAGTACCATCTGAGCGCGCATCTGACTCACAGAGAGAAAGCCGTTGTTG 480
Db 421 CAGGAGTACCATCTGAGCGCGCATCTGACTCACAGAGAGAAAGCCGTTGTTG 480
Qy 481 TGCAGCACTGGCTGTGATCAAAATTAACACAAATCAACTTGAAGAAATTTTGA 540
Db 481 TGCAGCACTGGCTGTGATCAAAATTAACACAAATCAACTTGAAGAAATTTTGA 540
Qy 541 ACGCAAAACATGAAATTAACAAAAATATATATATATATATATATATATATATATAT 600
Db 541 ACGCAAAACATGAAATTAACAAAAATATATATATATATATATATATATATATATAT 600
Qy 601 CTTTAAAGACATGACGAGTGAATTCATGAGTGCAGCATACCAATGAAGCTCTATT 660
Db 601 CTTTAAAGACATGACGAGTGAATTCATGAGTGCAGCATACCAATGAAGCTCTATT 660
Qy 661 CAAAGTGTACCCAGAGAGATGTGGAAACATTTGCATCACAGCAAGCTGAAAGACA 720
Db 661 CAAAGTGTACCCAGAGAGATGTGGAAACATTTGCATCACAGCAAGCTGAAAGACA 720
Qy 721 TCCCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTGTGGCAAAACATG 780
Db 721 TCCCAAGGCCACGAGGCTATGTATGTCAAAAAGATGTTCTTGTGGCAAAACATG 780
Qy 781 GACGGAATCTTGAAACATGTGAGAGAAACCCATTAAGAGAAATATCTATGGAAGTATG 840
Db 781 GACGGAATCTTGAAACATGTGAGAGAAACCCATTAAGAGAAATATCTATGGAAGTATG 840
Qy 841 CCGGAAACATTTAAACGCAAGATTAACCTTAAGCAACATGAAGAAATCATAGCCCGACA 900
Db 841 CCGGAAACATTTAAACGCAAGATTAACCTTAAGCAACATGAAGAAATCATAGCCCGACA 900
Qy 901 AAGCGATGTATGTGCTCTCCCAAGAGAGCTGTGGAAGAGCTATATCTATCTGTTTAA 960
Db 901 AAGCGATGTATGTGCTCTCCCAAGAGAGCTGTGGAAGAGCTATATCTATCTGTTTAA 960
Qy 961 TCTCCAAAGCCATATCTCTCTTCATGAGGAAGCGCCCTTTGTGTGAACATGC 1020
Db 961 TCTCCAAAGCCATATCTCTCTTCATGAGGAAGCGCCCTTTGTGTGAACATGC 1020
Qy 1021 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGCTCAGTACATGCTGTTGTACATGA 1080
Db 1021 TGGCTGTGGCAAAACATTTGCAATGAAGCAAGCTCAGTACATGCTGTTGTACATGA 1080
Qy 1081 TCCGTACAGAGAAATGAAGCTCAAAAGTCAAAATATCTGGAAGAAAGCGAGTTTGGC 1140
Db 1081 TCCGTACAGAGAAATGAAGCTCAAAAGTCAAAATATCTGGAAGAAAGCGAGTTTGGC 1140
Qy 1141 CTCCTATCTCAGTGTATATTCCTCCCAAGAGAAACAAAGGCAAGGCTTATCTTTGTG 1200
Db 1141 CTCCTATCTCAGTGTATATTCCTCCCAAGAGAAACAAAGGCAAGGCTTATCTTTGTG 1200
Qy 1201 TCAAAAGAGAGTCAACCACTGTGTGAAGCAAGATCTCTCAGAGTTGCAAGTACT 1260
Db 1201 TCAAAAGAGAGTCAACCACTGTGTGTGAAGCAAGATCTCTCAGAGTTGCAAGTACT 1260

Qy 1261 TACCCCTTGCTAA 1273
Db 1261 TACCCCTTGCTAA 1273
RESULT 4
US-09-831-426c-3
Sequence 3, Application US/09831426C
GENERAL INFORMATION:
APPLICANT: Hoechst Marion Roussel
APPLICANT: Borden-Pallier, F.
APPLICANT: Roher, C.
TITLE OF INVENTION: Human hUFI1A gene and coded hUFI1A protein
FILE REFERENCE: 146.1364
CURRENT APPLICATION NUMBER: US/09/831,426C
CURRENT FILING DATE: 2001-05-08
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Vers. 2.0
SEQ ID NO 3
LENGTH: 1273
TYPE: DNA
ORGANISM: Human
US-09-831-426c-3
Query Match 100.0%; Score 1273; DB 32; Length 1273;
Best Local Similarity 100.0%; Pred. No. 1.9e-279;
Matches 1273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCCAGCAGCGGCGGCGAGCGGCGGCTGCTGTGACGCGCGGCGCTCCGGAAGT 60
Db 1 ATGCCAGCAGCGGCGGCGGCGGCGGCTGCTGTGACGCGCGGCGCTCCGGAAGT 60
Qy 61 GTGCCGCGCTGCGGGAAGTTCAAGAGGAGCCGTGGCGCGCGCGCGCTCCGCGC 120
Db 61 GTGCCGCGCTGCGGGAAGTTCAAGAGGAGCCGTGGCGCGCGCGCGCTCCGCGC 120
Qy 121 ACCTGTCTGCGCAGCTGGCAGCGCGCTGCGCTGGGCTTGGAGGCGCGCGCGCTGGA 180
Db 121 ACCTGTCTGCGCAGCTGGCAGCGCGCTGCGCTGGGCTTGGAGGCGCGCGCGCTGGA 180
Qy 181 TCCGCGCGCGCTGGTGGCGGAGTGGTGGTCTTGGACATCCCGAGCGCTTCATTGGC 240
Db 181 TCCGCGCGCGCTGGTGGCGGAGTGGTGGTCTTGGACATCCCGAGCGCTTCATTGGC 240
Qy 241 AGCCGCGAGAGCTCAGCTCCGAGCCCGCGCGCGCGCTTCCGAGAGGTTTCATCTG 300
Db 241 AGCCGCGAGAGCTCAGCTCCGAGCCCGCGCGCGCGCTTCCGAGAGGTTTCATCTG 300
Qy 301 CTCCTTCCCTGACTGACGCGCAATTAAGCAAAAGCCTTGAAGCTTGAAGCTTGA 360
Db 301 CTCCTTCCCTGACTGACGCGCAATTAAGCAAAAGCCTTGAAGCTTGAAGCTTGA 360
Qy 361 CAAAGCAAGCGGCGAGAGACCATTTGTTGTGACTATGAAGGTTGGCAAGGCTTCAT 420
Db 361 CAAAGCAAGCGGCGAGAGACCATTTGTTGTGACTATGAAGGTTGGCAAGGCTTCAT 420
Qy 421 CAGGAGTACCATCTGAGCGCGCATCTGAGCTCACAGAGAGAAAGCGCTTGTG 480
Db 421 CAGGAGTACCATCTGAGCGCGCATCTGAGCTCACAGAGAGAAAGCGCTTGTG 480
Qy 481 TGCAGCACTGGCTGTGATCAAAATTAACACAAATCAACTTGAAGAAATTTTGA 540
Db 481 TGCAGCACTGGCTGTGATCAAAATTAACACAAATCAACTTGAAGAAATTTTGA 540
Qy 541 ACGCAAAACATGAAATTAACAAAAATATATATATATATATATATATATATATATAT 600
Db 541 ACGCAAAACATGAAATTAACAAAAATATATATATATATATATATATATATATATAT 600
Qy 601 CTTTAAAGACATGACGAGTGAATTCATGAGTGCAGCATACCAATGAAGCTCTATT 660
Db 601 CTTTAAAGACATGACGAGTGAATTCATGAGTGCAGCATACCAATGAAGCTCTATT 660

```
QY 661 CAAGGTACCCAGAGGATGTGGAAACACTTTCATGACCCAGCAACTGAACACA 720
    |||
Db 661 CAAGGTACCCAGAGGATGTGGAAACACTTTCATGACCCAGCAACTGAACACA 720
QY 721 TGCAGAGGCCACGAGGCTATGTATCTCAAAAAAGATGTCCTTGTGTGGCAAAACATG 780
    |||
Db 721 TGCAGAGGCCACGAGGCTATGTATCTCAAAAAAGATGTCCTTGTGTGGCAAAACATG 780
QY 781 GACGGAACCTCTGAAACACTGTGAGAAACCCATTAAGAGGAAATCTGTGAAGTATG 840
    |||
Db 781 GACGGAACCTCTGAAACACTGTGAGAAACCCATTAAGAGGAAATCTGTGAAGTATG 840
QY 841 CCGGAAAACTTTAAACGCAAAAGATTACCTTAAGCAACATGAAAACTCATGCCCCAGA 900
    |||
Db 841 CCGGAAAACTTTAAACGCAAAAGATTACCTTAAGCAACATGAAAACTCATGCCCCAGA 900
QY 901 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGACCTATACACTGTGTAA 960
    |||
Db 901 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGACCTATACACTGTGTAA 960
QY 961 TCTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGC 1020
    |||
Db 961 TCTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGC 1020
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTACGATGCTGTACATGA 1080
    |||
Db 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTACGATGCTGTACATGA 1080
QY 1081 TCCTGACAGAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGA AAAACGAGTTTGGC 1140
    |||
Db 1081 TCCTGACAGAGAAATGAAGCTCAAAAGTCAAAAATCTCGTGA AAAACGAGTTTGGC 1140
QY 1141 CTCCTATCTCAGTGTATATCCCTCCCAAAAGAAACAAAGGCAAGGCTTATCTTTTG 1200
    |||
Db 1141 CTCCTATCTCAGTGTATATCCCTCCCAAAAGAAACAAAGGCAAGGCTTATCTTTTG 1200
QY 1201 TCAAAACGAGAGTACCCCAACTGTGTGAAGACAAGATCCTGACAGTTGCACTACT 1260
    |||
Db 1201 TCAAAACGAGAGTACCCCAACTGTGTGAAGACAAGATCCTGACAGTTGCACTACT 1260
QY 1261 TACCTTGGCTAA 1273
    |||
Db 1261 TACCTTGGCTAA 1273

RESULT 5
US-09-644-868-9602
; Sequence 9602, Application US/09644868
; GENERAL INFORMATION:
; APPLICANT: White, David
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: 1600,1168-001
; CURRENT APPLICATION NUMBER: US/09/644,868
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 60/151,063
; NUMBER OF SEQ ID NOS: 10075
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9602
; LENGTH: 1597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(1597)
; OTHER INFORMATION: n = A,T,C or G
US-09-644-868-9602

Query Match 99.7%; Score 1269.8; DB 25; Length 1597;
Best Local Similarity 99.8%; Pred. No. 1,1e-278;
```

```
Matches 1271: Conservative 0: Mismatches 2: Indels 0: Gaps 0:
QY 1 ATGCCAGACAGCGGCGGACCGGCGGCTGCTGTGACCGCGCGCTCCGGAAGT 60
    |||
Db 18 ATGCCAGACAGCGGCGGCGGACCGGCGGCTGCTGTGACCGCGCGCTCCGGAAGT 77
QY 61 GTGCCGCGGTGCGGGAAGGTTCAGCAGGAGAGCGTGTGGCCGCGGCGGCTTCCGCGC 120
    |||
Db 78 GTGCCGCGGTGCGGGAAGGTTCAGCAGGAGAGCGTGTGGCCGCGGCGGCTTCCGCGC 137
QY 121 ACCTGTCTGGCAGCTGGCAAGCGGCTGGCCCTTGGGCTTGAAGCGCGCGGCTTGGCA 180
    |||
Db 138 ACCTGTCTGGCAGCTGGCAAGCGGCTGGCCCTTGGGCTTGGGAGCGCGGCGGCTTGGCA 197
QY 181 TCCGCGCGCGGTGGTGGCGCGAGTGGTGTGCTGCTGACATTCGCGAGCGCTTCAATTG 240
    |||
Db 198 TCCGCGCGCGGTGGTGGCGCGAGTGGTGTGCTGCTGACATTCGCGAGCGCTTCAATTG 257
QY 241 AACCAGGAGAGCTGAGTCCGACCCCGCGCGCCGCGCTTCCAGAGGTTTCATCTG 300
    |||
Db 258 AACCAGGAGAGCTGAGTCCGACCCCGCGCGCCGCGCTTCCAGAGGTTTCATCTG 317
QY 301 CTCCTTCCCTGACTGACGCGCAATTCACGAAAGCTGGAAGCTTGACGCGCACCTGTG 360
    |||
Db 318 CTCCTTCCCTGACTGACGCGCAATTCACGAAAGCTTGAAGCTTGACGCGCACCTGTG 377
QY 361 CAAGCACACGCGGCGAGAGCCATTTGTTGTGACTATGAAGGCTGTGGCAAGCCCTTCAT 420
    |||
Db 378 CAAGCACACGCGGCGAGAGCCATTTGTTGTGACTATGAAGGCTGTGGCAAGCCCTTCAT 437
QY 421 CAGGGACACACATCTGACCGCGCACATTTGACCTGACACAGAGGAAACCGTTGTTTG 480
    |||
Db 438 CAGGGACACACATCTGACCGCGCACATTTGACCTGACACAGAGGAAACCGTTGTTTG 497
QY 481 TCCAGCCACTGCTGTGTATCAAAAATTCACACAAAATCAACTGGAAGAAACATTTTGA 540
    |||
Db 498 TCCAGCCACTGCTGTGTATCAAAAATTCACACAAAATCAACTGGAAGAAACATTTTGA 557
QY 541 ACGCAACATGAAATTAACAAACAAATATATATGACAGTTTGAAGACTGTAAACAGC 600
    |||
Db 558 ACGCAACATGAAATTAACAAACAAATATATATGACAGTTTGAAGACTGTAAACAGC 617
QY 601 CTTTAAAGAACTACGACGCTGAAATTCATCATGTGCGACATACCATGAACCTCTATT 660
    |||
Db 618 CTTTAAAGAACTACGACGCTGAAATTCATCATGTGCGACATACCATGAACCTCTATT 677
QY 661 CAAGGTACCCAGAGGATGTGGAAACACTTTCATGACCCAGCAACTGAACACA 720
    |||
Db 678 CAAGGTACCCAGAGGATGTGGAAACACTTTCATGACCCAGCAACTGAACACA 737
QY 721 TGCAGAGGCCACGAGGCTATGTATCTCAAAAAAGATGTCCTTGTGTGGCAAAACATG 780
    |||
Db 738 TGCAGAGGCCACGAGGCTATGTATCTCAAAAAAGATGTCCTTGTGTGGCAAAACATG 797
QY 781 GACGGAACCTCTGAAACACTGTGAGAAACCCATTAAGAGGAAATCTGTGAAGTATG 840
    |||
Db 798 GACGGAACCTCTGAAACACTGTGAGAAACCCATTAAGAGGAAATCTGTGAAGTATG 857
QY 841 CCGGAAAACTTTAAACGCAAAAGATTACCTTAAGCAACATGAAAACTCATGCCCCAGA 900
    |||
Db 858 CCGGAAAACTTTAAACGCAAAAGATTACCTTAAGCAACATGAAAACTCATGCCCCAGA 917
QY 901 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGACCTATACACTGTGTAA 960
    |||
Db 918 AAGGATGTATGTGCTGTCCAGAGAAAGCTGTGAAAGACCTATACACTGTGTAA 977
QY 961 TCTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGC 1020
    |||
Db 978 TCTCCAAAGCCATATCTCTCTTCATGAGAAAGCCGCTTTGTGTGAACATGC 1037
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTACGATGCTGTGTACATGA 1080
    |||
Db 1038 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCAGTACGATGCTGTGTACATGA 1097
```

QY	1081	TCCGACAAACAAATAATACCTCAAGTCAAAATATCTGTCAAAAACGAGATTGGC	1140
Db	1098 <th>TCTGACAGAGAAATAATACCTCAAAAGTCAAAATATCTGTCAAAAACGAGATTGGC</th> <th>1157</th>	TCTGACAGAGAAATAATACCTCAAAAGTCAAAATATCTGTCAAAAACGAGATTGGC	1157
QY	1141 <th>CTCTCATCTCAGTGATATATCCCTCCCAAAAGCAACAGGACGCTTATCTTTGTG</th> <th>1200</th>	CTCTCATCTCAGTGATATATCCCTCCCAAAAGCAACAGGACGCTTATCTTTGTG	1200
Db	1158 <th>CTCTCATCTCAGTGATATATCCCTCCCAAAAGCAACAGGACGCTTATCTTTGTG</th> <th>1217</th>	CTCTCATCTCAGTGATATATCCCTCCCAAAAGCAACAGGACGCTTATCTTTGTG	1217
QY	1201 <th>TCAAAACGAGAGTACACCAACTGTGTGGAAGACAAAGATGTCTCGACAGTTCAGACT</th> <th>1260</th>	TCAAAACGAGAGTACACCAACTGTGTGGAAGACAAAGATGTCTCGACAGTTCAGACT	1260
Db	1218 <th>TCAAAACGAGAGTACACCAACTGTGTGGAAGACAAAGATGTCTCGACAGTTCAGACT</th> <th>1277</th>	TCAAAACGAGAGTACACCAACTGTGTGGAAGACAAAGATGTCTCGACAGTTCAGACT	1277
QY	1261 <th>TACCCTTGAGCTAA 1273</th> <td></td>	TACCCTTGAGCTAA 1273	
Db	1278 <th>TACCCTTGAGCTAA 1290</th> <td></td>	TACCCTTGAGCTAA 1290	

RESULT 6

US-09-552-126-8862

Sequence 8862, Application US/09652126

GENERAL INFORMATION:

APPLICANT: Shyban, Andrew W.

TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES

TITLE OF INVENTION: THEREFOR

FILE REFERENCE: 1600.1185-001

CURRENT APPLICATION NUMBER: US/09/652.126

CURRENT FILING DATE: 2000-08-30

PRIOR APPLICATION NUMBER: 60/151,132

PRIOR FILING DATE: 1999-08-30

NUMBER OF SEQ. ID NOS: 10051

SOFTWARE: FastSeq for Windows Version 4.0

SEQ. ID NO. 8862

LENGTH: 1597

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(1597)

OTHER INFORMATION: n - A,T,C or G

US-09-552-126-8862

Query Match

Best Local Similarity 99.7%; Score 1269.8; DB 25; Length 1597;

Matches 1271; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	ATGCGCAGCAGCGCGCCGACGCGGGGCGGTGCTGTGACCGCGCGCTCCCGAGAGT	60
Db	18 <th>ATGCGCAGCAGCGCGCCGACGCGGGGCGGTGCTGTGACCGCGCGCTCCCGAGAGT</th> <th>77</th>	ATGCGCAGCAGCGCGCCGACGCGGGGCGGTGCTGTGACCGCGCGCTCCCGAGAGT	77
QY	61 <th>GTCGCGGCGCGCGCCGACGCGGTTCACAGGAGACCGTGGGCGGGCGCGCGGTCCCGC</th> <th>120</th>	GTCGCGGCGCGCGCCGACGCGGTTCACAGGAGACCGTGGGCGGGCGCGCGGTCCCGC	120
Db	78 <th>GTCGCGGCGCGCGCCGACGCGGTTCACAGGAGACCGTGGGCGGGCGCGCGGTCCCGC</th> <th>137</th>	GTCGCGGCGCGCGCCGACGCGGTTCACAGGAGACCGTGGGCGGGCGCGCGGTCCCGC	137
QY	121 <th>ACGTGTCTCGGCACGTGCAGCGCCGCTTGGGCTTGGAGAGCCCGCGCCCTCTGA</th> <th>180</th>	ACGTGTCTCGGCACGTGCAGCGCCGCTTGGGCTTGGAGAGCCCGCGCCCTCTGA	180
Db	138 <th>ACGTGTCTCGGCACGTGCAGCGCCGCTTGGGCTTGGAGAGCCCGCGCCCTCTGA</th> <th>197</th>	ACGTGTCTCGGCACGTGCAGCGCCGCTTGGGCTTGGAGAGCCCGCGCCCTCTGA	197
QY	181 <th>TCCGCGGCGCGGTGTGCGCGAGTGGGTGCTGCTTACCACTATGCGCGACCGCTTCAATTGC</th> <th>240</th>	TCCGCGGCGCGGTGTGCGCGAGTGGGTGCTGCTTACCACTATGCGCGACCGCTTCAATTGC	240
Db	198 <th>TCCGCGGCGCGGTGTGCGCGAGTGGGTGCTGCTTACCACTATGCGCGACCGCTTCAATTGC</th> <th>257</th>	TCCGCGGCGCGGTGTGCGCGAGTGGGTGCTGCTTACCACTATGCGCGACCGCTTCAATTGC	257
QY	241 <th>AGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCCGCGCTTCCACGAGGTTTCATCTG</th> <th>300</th>	AGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCCGCGCTTCCACGAGGTTTCATCTG	300
Db	258 <th>AGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCCGCGCTTCCACGAGGTTTCATCTG</th> <th>317</th>	AGCGCGGAGAGCTCAGCTCCGACCCCGCGCGCCGCGCTTCCACGAGGTTTCATCTG	317
QY	301 <th>CTCTCTTCCGAGATGAGCGGCAATTACAGCAAAAGCGTGAAGGATTCAGCGGACCTTGG</th> <th>360</th>	CTCTCTTCCGAGATGAGCGGCAATTACAGCAAAAGCGTGAAGGATTCAGCGGACCTTGG	360
Db	318 <th>CTCTCTTCCGAGATGAGCGGCAATTACAGCAAAAGCTTGAAGGATTCAGCGGACCTTGG</th> <th>377</th>	CTCTCTTCCGAGATGAGCGGCAATTACAGCAAAAGCTTGAAGGATTCAGCGGACCTTGG	377
QY	361 <th>CAAGCACACGCGGAGAGACATTTCTTTGTGTACTATGAACGCTGTGGCAAGGCTTCAT</th> <th>420</th>	CAAGCACACGCGGAGAGACATTTCTTTGTGTACTATGAACGCTGTGGCAAGGCTTCAT	420
Db	378 <th>CAAGCACACGCGGAGAGACATTTCTTTGTGTACTATGAAGGCTGTGGCAAGGCTTCAT</th> <th>437</th>	CAAGCACACGCGGAGAGACATTTCTTTGTGTACTATGAAGGCTGTGGCAAGGCTTCAT	437

Qy	421	CAGGAGCTGCATGTCGAGCCGACATCTTCGACTGCACACGAGGAAAGCCGTTTGTG	489
Db	438	CAGGAGCTACATCTGAGCCGACATTTCTGACTGCACACGAGGAAAGCCGTTTGTG	497
Qy	481	TGCGACCTAGCTGCTGTGATCAAAAAATTCACACAAAAATCAAACTTGAAGAAACATTTTGA	540
Db	498	TGCGACCTAGCTGCTGTGATCAAAAAATTCACACAAAAATCAAACTTGAAGAAACATTTTGA	557
Qy	541	ACGGAACATGAAATTCACAAAAACAATATATATGACGTTTGAAGACTGAAGAGAC	600
Db	558	ACGGAACATGAAATTCACAAAAACAATATATATGACGTTTGAAGACTGAAGAGAC	617
Qy	601	CTTTAAGAAACATCAGCAGCTTGAAATTCATCAGTGCAGCATACCAATGAAACCTTAAT	660
Db	618	CTTTAAGAAACATCAGCAGCTTGAAATTCATCAGTGCAGCATACCAATGAAACCTTAAT	677
Qy	661	CAAGTATACCCAGGAAGATGTGGGAAACCTTGCATACCCAGCAGCTGAAACACA	720
Db	678	CAAGTATACCCAGGAAGATGTGGGAAACCTTGCATACCCAGCAGCTGAAACACA	737
Qy	721	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAAGATGTTCTTTGTGGCAAAAACATG	780
Db	738	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAAGATGTTCTTTGTGGCAAAAACATG	797
Qy	781	GAGCGAATCTTGAAACATGTGAGAGAAACCCATAAGAGAAATACTATGTGAAGATG	840
Db	798	GAGCGAATCTTGAAACATGTGAGAGAAACCCATAAGAGAAATACTATGTGAAGATG	857
Qy	841	CCGGAACATCTTAAACCAAGATTAAGCAACATGAAACCTCATGCCCCAGA	900
Db	858	CCGGAACATCTTAAACCAAGATTAAGCAACATGAAACCTCATGCCCCAGA	917
Qy	901	AAGGATGTATGTCCGTGCTCCAGAGAAAGCTGTGAGAACCTTACTACTGTGTTAA	960
Db	918	AAGGATGTATGTCCGTGCTCCAGAGAAAGCTGTGAGAACCTTACTACTGTGTTAA	977
Qy	961	TCTCCAAAGCCATATCTCTCTCCATGAGGAAAGCCGCTTTGTGTGAACATGC	1020
Db	978	TCTCCAAAGCCATATCTCTCTCCATGAGGAAAGCCGCTTTGTGTGAACATGC	1030
Qy	1021	TGGCTGTGGCAAAACATTTGCAATGAANCAAACTCTCCTAGAGCAATGCTGTGTAACATGA	1080
Db	1038	TGGCTGTGGCAAAACATTTGCAATGAANCAAACTCTCCTAGAGCAATGCTGTGTAACATGA	1099
Qy	1081	TCCTGCACAGAGAAATATGAAAGCTCAAAAGTCAAAAAATCTCGTGAAAAAACGAGATTGGC	1144
Db	1098	TCCTGCACAGAGAAATATGAAAGCTCAAAAGTCAAAAAATCTCGTGAAAAAACGAGATTGGC	1151
Qy	1141	CTCTCATCTCAGAGATATATCCCTCCCAAAAGGAAACAAAGGCAAGGCTTATCTTTGTG	1200
Db	1158	CTCTCATCTCAGAGATATATCCCTCCCAAAAGGAAACAAAGGCAAGGCTTATCTTTGTG	1211
Qy	1201	TCAAAACGAGAGTCAACCACTGTGTGAAGACACAGATGCTCTGACAGTTTGCACTACT	1266
Db	1218	TCAAAACGAGAGTCAACCACTGTGTGAAGACACAGATGCTCTGACAGTTTGCACTACT	1277
Qy	1261	TACCTTGGCTAA 1273	
Db	1278	TACCTTGGCTAA 1290	
RESULT 7			
US-09-652-814-10703			
; Sequence 10703, Application us/09652814			
; GENERAL INFORMATION:			
; APPLICANT: Holtzman, Douglas A.			
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
; TITLE OF INVENTION: THEREFOR			
; FILE REFERENCE: 1600.1191-001			
; CURRENT APPLICATION NUMBER: US/09/652, 814			
; CURRENT FILING DATE: 2000-08-31			
; PRIORITY APPLICATION NUMBER: 60/152, 109			

Q7	181	TCCGCGCGCCGCTGTCGTCGCCAGATGGGTGTGTCTCTTGACCATCGCCGACGGCTTCATTGC	240
Db	198	TCCGCGCGCCGCTGTCGTCGCCAGATGGGTGTGTCTCTTGACCATCGCCGACGGCTTCATTGC	257
Q7	241	AGCGCGCGAGAGCTCAGCTCCGACCCCGCGCCGCCGCGCTTCCGAGAGGTTCACTCTG	300
Db	258	AGCGCGCGAGAGCTCAGCTCCGACCCCGCGCCGCCGCGCTTCCGAGAGGTTCACTCTG	317
Q7	301	CTCCCTTCCCTGACTGCAGCGCCATTACACAAAGCCTTGAAGCTTGACGGGCACTGTG	360
Db	318	CTCCCTTCCCTGACTGCAGCGCCATTACACAAAGCCTTGAAGCTTGACGGGCACTGTG	377
Q7	361	CAAGCACAGCGGGGGAGAGACCATTTGTTTGTGACATGAAGGGTTGGCAGAGGCTTCAT	420
Db	378	CAAGCACAGCGGGGGAGAGACCATTTGTTTGTGACATGAAGGGTTGGCAGAGGCTTCAT	437
Q7	421	CAGGGACTACATCTGAGCGCCGACATTTCTGACTCCACAGAGAGAAAAGCCGTTTGTG	480
Db	438	CAGGGACTACATCTGAGCGCCGACATTTCTGACTCCACAGAGAGAAAAGCCGTTTGTG	497
Q7	481	TGCAGCCACTGGCTGTGATCAAAAATTCAACACACAAATCAAACTTGAAGAAACATTTCG	540
Db	498	TGCAGCCATGGCTGTGATCAAAAATTCAACACACAAATCAAACTTGAAGAAACATTTCG	557
Q7	541	AGCGAAACATGAAAATTCACAAAAACATATATATCAGTTTGAAGACTTAAAGAGAC	600
Db	558	AGCGAAACATGAAAATTCACAAAAACATATATATCAGTTTGAAGACTTAAAGAGAC	617

```

OY      1261  TACCGTTGGCTAA 1273
        |||
Db      1278  TACCGTTGGCTAA 1290

RESULT 9
US-09-716-990-1322
; Sequence 1322, Application US/09716990
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2039-001
; CURRENT APPLICATION NUMBER: US/09/716,990
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: US 60/166,926
; PRIOR FILING DATE: 1999-11-22
; NUMBER OF SEQ. ID NOS: 1691
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1322
; LENGTH: 1597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(157)
; OTHER INFORMATION: n = A,T,C or G
US-09-716-990-1322

```

Qy	601	CTTAAAGAAACATCAGACGCGTGAATATCATCATGCTCCAGCATACCATATGAACCTCATTT	660
Db	618	CTTTAAAGAAACATCAGACGCGTGAATATCATCATGCTCCAGCATACCATATGAACCTCATTT	677
Qy	661	CAAGTGTACCCAGGAAGGATGTGGGAAACACTTTGGCATACCCAGCAAGCTGAAAGCACA	720
Db	678	CAAGTGTACCCAGGAAGGATGTGGGAAACACTTTGGCATACCCAGCAAGCTGAAAGCACA	737
Qy	721	TGCGAAGGCCCGACGAGGGCTGTGTATGTGTCAAAAAGGATGTCTCTTGTGGCAAAAAGATG	780
Db	738	TGCGAAGGCCCGACGAGGGCTGTGTATGTGTCAAAAAGGATGTCTCTTGTGGCAAAAAGATG	797
Qy	781	GACGGAACCTTCTGAACACTGTGAGAGAAACCCATTAAGAGAAATATATGTGTGAAGTATG	840
Db	798	GACGGAACCTTCTGAACACTGTGAGAGAAACCCATTAAGAGAAATATATGTGTGAAGTATG	857
Qy	841	CCGGAAACATTTTAAACCCCAAGATTTACTCTTAAGCAACACATGAAGAACTCATGTGCCCGA	900
Db	858	CCGGAAACATTTTAAACCCCAAGATTTACTCTTAAGCAACACATGAAGAACTCATGTGCCCGA	917
Qy	901	AAGGATGTATGTGCTGCTGCCAAGAGAAAGCGCTGSGAAGAACCTTACTACTGTGTTTAA	960
Db	918	AAGGATGTATGTGCTGCTGCCAAGAGAAAGCGCTGSGAAGAACCTTACTACTGTGTTTAA	977
Qy	961	TCTCCAAAGCCATATCTCTCTCTTCCATGTAGAGAAACCCGCCCTTTTGTGTGAACATATG	1020
Db	978	TCTCCAAAGCCATATCTCTCTCTTCCATGTAGAGAAACCCGCCCTTTTGTGTGAACATATG	1033
Qy	1021	TGGCTGTGGCAAAACATTTTGAATGAATGAACAAAGCTACATAGSCATGCTGTGTATGACATGA	1088
Db	1038	TGGCTGTGGCAAAACATTTTGAATGAATGAACAAAGCTACATAGSCATGCTGTGTATGACATGA	1099
Qy	1081	TCTGTACAAAGAAATATGAAGCTCAAAAGTCAAAAAATCTGTGAAAAAAGGAGTTTGGC	1144
Db	1098	TCTGTACAAAGAAATATGAAGCTCAAAAGTCAAAAAATCTGTGAAAAAAGGAGTTTGGC	1155
Qy	1141	CTCTCATCTCAGTGTATATATCCCTCCCAAAAGGAAACAAGGCGCAAGCTATCTTTGTG	1200
Db	1158	CTCTCATCTCAGTGTATATATCCCTCCCAAAAGGAAACAAGGCGCAAGCTATCTTTGTG	1211
Qy	1201	TCAAAAGGAGAGTCAACCCCACTGTGTGGAGAAACAAGATGTCTGTGACATGTGGACATCT	1266
Db	1218	TCAAAAGGAGAGTCAACCCCACTGTGTGGAGAAACAAGATGTCTGTGACATGTGGACATCT	1273
Qy	1261	TACCTTGGCTTAA	1273
Db	1278	TACCTTGGCTTAA	1290

```

RESULT 10
? Sequence 6336, Application US/09721589
? GENERAL INFORMATION:
? APPLICANT: Gearing, David P.
? APPLICANT: Villeva, Jean-Luc
? TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
? TITLE OF INVENTION: THEREFOR
? FILE REFERENCE: 1600.2045-001
? CURRENT APPLICATION NUMBER: US/09/721,589
? CURRENT FILING DATE: 2000-11-22
? PRIOR APPLICATION NUMBER: 60/167,380
? PRIOR FILING DATE: 1999-11-24
? NUMBER OF SEQ ID NOS: 7017
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 6336
? LENGTH: 1597
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)..(1597)
? OTHER INFORMATION: n = A,T,C or G

```

[illegible]

Qy	1021	TGGCTGTGGCAAAACATTGGCATGAAACAAAGTCTCACTAGGATGCTCTGTGTACATGA	1080
Db	1038	TGGCTGTGGCAAAACATTGGCATGAAACAAAGTCTCACTAGGATGCTCTGTGTACATGA	1097
Qy	1081	TCTGCAAGAGAAATGAGAGCTCAAGTCAAAAAATCTGTGTAACGAGCTTTGGC	1140
Db	1098	TCTGCAAGAGAAATGAGAGCTCAAGTCAAAAAATCTGTGTAACGAGCTTTGGC	1157
Qy	1141	CTCTCATCTCACTGATGTATTCCTGCCAAAGAGAACAGGCGTTATCTTTGTG	1200
Db	1158	CTCTCATCTCACTGATGTATTCCTGCCAAAGAGAACAGGCGTTATCTTTGTG	1217
Qy	1201	TCAAAAGGAGAGTACCCCACTGTGTGGAGACAGATGCTCTCGACAGTTGCAGTACT	1260
Db	1218	TCAAAAGGAGAGTACCCCACTGTGTGGAGACAGATGCTCTCGACAGTTGCAGTACT	1277
Qy	1261	TACCTTGGCTAA 1273	
Db	1278	TACCTTGGCTAA 1290	

```

RESULT 1
US-09-726-788-6896
: Sequence 6896, Application US/09726788
: GENERAL INFORMATION:
: APPLICANT: Gearing, David P.
: APPLICANT: Kingsbury, Gillian A.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MO
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.2051-001
: CURRENT APPLICATION NUMBER: US/09/726,788
: CURRENT FILING DATE: 2000-11-30
: PRIOR APPLICATION NUMBER: 60/168,131
: PRIOR FILING DATE: 1999-11-30
: NUMBER OF SEQ ID NOS: 7691
: SOFTWARE: FASTEST for Windows Version 4.0.
: SEQ ID NO 6896
: LENGTH: 1597
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(1597)
: OTHER INFORMATION: n = A,T,C or G
US-09-726-788-6896

```

Query Match	99.78;	Score 1269.8;	DB 29;	Length 1597;
Best Local Similarity	99.88;	Pred. No. 1.1e-278;		
Matches 1271; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

QY	1	ATGCGAGCAGCGGCGCCGACGCGGGCGGTGCTGTTGACGCGCGGCGCTCCGCGAAGT	60
Db	18	ATGCGAGCAGCGGCGCGCGACGCGGGCGGTGCTGTTGACGCGCGGCGCTCCGCGAAGT	77
QY	61	GTCGCGGCGCTCCGCGCGAGAGTTTCAGCAGAGCGCGTGGCGCGGCGCGCGCGCTTCCGGCG	120
Db	78	GTCGCGGCGCTCCGCGCGAGAGTTTCAGCAGAGCGCGTGGCGCGGCGCGCGCGCTTCCGGCG	137
QY	121	ACGTGTCGTGCGCAGCTGTCGACGCGCGCCCTGCGCGGCGCTTGGAGGCGCGCGCGCCCTGGA	180
Db	138	ACGTGTCGTGCGCAGCTGTCGACGCGCGCCCTGCGCGGCGCTTGGAGGCGCGCGCGCCCTGGA	197
QY	181	TCCGCGCGCGCTGTGTCGCGCGAGTCGATGTCGCTTCGACCAATGCGCGACGCCCTTCATTGC	240
Db	198	TCCGCGCGCGCTGTGTCGCGCGAGTCGATGTCGCTTCGACCAATGCGCGACGCCCTTCATTGC	257
QY	241	AGCGCGCGAGAGCTCATGCTCCGACGCGCGCGCGCGCGCGCGCTTCCAGAGAGTTCACTCG	300
Db	258	AGCGCGCGAGAGCTCATGCTCCGACGCGCGCGCGCGCGCGCTTCCAGAGAGTTCACTCG	317
QY	301	CTCTCTTCCCTGACCTCAGCCGCAATTACAGCAAAAGCCTGAGAGCTTGACGGGCACTGTG	360

Db	318	CTCTGTTCCCTGACGTGACGGCCANTTTACACCAAAAGCGTTGGAAGCTTGAACGGGCACACTGTG	3177
Qy	361	CAACACACAGGGGGAGACACCAATTTGTTTGTGACTATGAAGCGTGTGCCAAGCCCTTCAT	420
Db	378	CAAGCACACGGGGGAGAGACCAATTTGTTTGTGACTATGAAGCGTGTGCCAAGCGCTTCAT	437
Qy	421	CAGGACACTGCATCTGAGCGGCCACATTTCTGACTCACAGGAGAAAGAGCCGTTTGTTCG	480
Db	438	CAGGACACTGCATCTGAGCGGCCACATTTCTGACTCACAGGAGAAAGAGCCGTTTGTTCG	497
Qy	481	TGCAGCCACTGCGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAAGAAACATTTTGA	540
Db	498	TGCAGCCCATGCGCTGTGATCAAAAATTTCAACACAAAATCAAACTTGAAGAAACATTTTGA	557
Qy	541	ACGCAAAACATGAAAATTCAAACAAAACAAATATATGACGTTTGAAGACTGTAAAGAC	600
Db	558	ACGCAAAACATGAAAATTCAAACAAAACAAATATATGACGTTTGAAGACTGTAAAGAAC	617
Qy	601	CTTTAAGAAACATCAGCAGGCTGAATTCATCAGTCCAGCATACCAATGAACCTGTATT	660
Db	618	CTTTAAGAAACATCAGCAGGCTGAATTCATCAGTCCAGCATACCAATGAACCTGTATT	677
Qy	661	CAAGTACCCAGGAAAGGATGGGAAACCTTTGATCAGCCAGCAGCTGAAGACACA	720
Db	678	CAAGTATCCAGGAAAGGATGGGAAACCTTTGATCAGCCAGCAGCTGAAGACACA	737
Qy	721	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATG	780
Db	738	TGCCAAGGCCACAGAGGCTATGTATGTCAAAAAGCATGTTCTTTGTGGCAAAAACATG	797
Qy	781	GAGGGAACCTTCTGAAACATGTGAGAGAAACCCATTAAGAGGAAATATCTATGTGAAGATG	840
Db	798	GAGGGAACCTTCTGAAACATGTGAGAGAAACCCATTAAGAGGAAATATCTATGTGAAGATG	857
Qy	841	CCGGAAAACATTTAAACCGCAAAAGATTACCTTAAGCAACATGAAGAACTCATGGCCGACA	900
Db	858	CCGGAAAACATTTAAACCGCAAAAGATTACCTTAAGCAACATGAAGAACTCATGGCCGACA	917
Qy	901	AAGGAGATGATGCGCTGTGCCAAGAAAGCGCTGTGAAGAACCTTACTACTGTGTTTAA	960
Db	918	AAGGAGATGATGCGCTGTGCCAAGAAAGCGCTGTGAAGAACCTTACTACTGTGTTTAA	977
Qy	961	TCTCCAAAGCATATCTCTCTCTTCATGAGGAAGCGCCCTTTGTGTGTGAACATGC	1020
Db	978	TCTCCAAAGCATATCTCTCTCTTCATGAGGAAGCGCCCTTTGTGTGTGAACATGC	1030
Qy	1021	TGGCTGTGGCAAAACATTTGCAATGAAGAAACAAAGTCTCAGTACGATGCTGTGTATCATGA	1088
Db	1038	TGGCTGTGGCAAAACATTTGCAATGAAGAAACAAAGTCTCAGTACGATGCTGTGTATCATGA	1099
Qy	1081	TCTCTGACAAAGAAATGAAGGCTCAAAAGTCAAAAATCTCTGTGAAGAAACGAGCTTGGC	1144
Db	1098	TCTCTGACAAAGAAATGAAGGCTCAAAAGTCAAAAATCTCTGTGAAGAAACGAGCTTGGC	1155
Qy	1141	CTCTCATCTTCATGATATATTCCTCCCAAAAGAAACAAGGCAAGCTTATCTTTGTG	1200
Db	1158	CTCTCATCTTCATGATATATTCCTCCCAAAAGAAACAAGGCAAGCTTATCTTTGTG	1211
Qy	1201	TCAAAACGAGAGATCACCCCAACTGTGTGGAAGACAAAGATGCTCTGACAGTTGCAGTACT	1261
Db	1218	TCAAAACGAGAGATCACCCCAACTGTGTGGAAGACAAAGATGCTCTGACAGTTGCAGTACT	1277
Qy	1261	TACCTTTGGCTAA	1273
Db	1278	TACCTTTGGCTAA	1290

RESULT 12
US-60-324-185-16829
; Sequence 16829, Application US/60324185
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti


```
Db 20 ATGCCGACGACGCGCGCGACGCGGGCGGTGCTGCTGACCGCGCGCTCCCGAAGT 79
QY 61 GTGCGCGCGCTCGCGCGAAGGTTTCAGCAGGAGCGCGTGGCGCGCGCGCGCTCCCGC 120
Db 80 GTGCGCGCGCGCGCGCGAAGGTTTCAGCAGGAGCGCGTGGCGCGCGCGCGCGCTCCCGC 139
QY 121 ACCTGTCTCGGACGCT-GGCAGCGCGCGCTGGCGCT-GGGCTTGGAGGCGCGCGCGCTG 178
Db 140 ACCTGTCTCGGACGCTGGCGAGCGCGCTGGCGCTGGCGCTGGAGGCGCGCGCGCTG 199
QY 179 GATCGCGCGCGCGCGTGGTGGCGGAGTGGGTGCTGCTTGCATCGCGCGAGCGCTT-CAT 237
Db 200 GATCGCGCGCGCGCGTGGTGGCGGAGTGGGTGCTGCTTGCATCGCGCGAGCGCTTGCAT 259
QY 238 TGCAGCGCGCGAGAGCTTCCGACCGCGCGCGCGCGCGC-TTCCAGAGAGTTCA 296
Db 260 TGCAGCGCGCGAGAGCTTCCGACCGCGCGCGCGCGCGCTTCCAGAGAGTTCA 319
QY 297 TCTGCTCCTTCCCTGACTGCGCGCGCAATTACAGCAAAAGCTTGGAGCGCGCAC 356
Db 320 TCTGCTCCTTCCCTGACTGCGCGCGCAATTACAGCAAAAGCTTGGAGCGCGCAC 379
QY 357 TGTGCAACACACGCGGAGAGAGACATTTGTTGTGACTATGAAAGGTTGGCAAGGCT 416
Db 380 TGTGCAACACACGCGGAGAGAGACATTTGTTGTGACTATGAAAGGTTGGCAAGGCT 439
QY 417 TCATCAGGACTACCATCTTGAGCGCGCACTTGTACTCACAGAGAGAAACCGTTTG 476
Db 440 TCATCAGGACTACCATCTTGAGCGCGCACTTGTACTCACAGAGAGAAACCGTTTG 499
QY 477 TTTGTGACGCGCATGCTGCTGATCAAAAATTCAACAAATCAACTGGAAGAACTT 536
Db 500 TTTGTGACGCGCAATGGCTGTGATCAAAAATTCAACAAATCAACTGGAAGAACTT 559
QY 537 TTGAACGCAAAACATGAAATATACAAAACATATATATGACATTTTGAAGCTGTAGA 596
Db 560 TTGAACGCAAAACATGAAATATACAAAACATATATATGACATTTTGAAGCTGTAGA 619
QY 597 AGACCTTTAAGAAACATGACAGCTGAAATTCATCAGTGCAGCATCAATGAACTG 656
Db 620 AGACCTTTAAGAAACATGACAGCTGAAATTCATCAGTGCAGCATCAATGAACTG 679
QY 657 TATTCAGTGTACCGAGAGAGATGTGGGAAACATTTGATCAGCGACGAAGCTGAAC 716
Db 680 TATTCAGTGTACCGAGAGAGATGTGGGAAACATTTGATCAGCGACGAAGCTGAAC 739
QY 717 GACATGCGCAAGCGCCAGAGGCTATGTATGCAAAAAGATGTTCTTGTGGCAAAA 776
Db 740 GACATGCGCAAGCGCCAGAGGCTATGTATGCAAAAAGATGTTCTTGTGGCAAAA 799
QY 777 CATGAGCGGAATCTTGTGAACATGTGAGAGAAACCATTAAGAGGAAATACATATGAG 836
Db 800 CATGAGCGGAATCTTGTGAACATGTGAGAGAAACCATTAAGAGGAAATACATATGAG 859
QY 837 TATGCGCGGAAAAATTTAAACGCAAGATTTACCTTAAGCAGACATGAAACATCATGCGC 896
Db 860 TATGCGCGGAAAAATTTAAACGCAAGATTTACCTTAAGCAGACATGAAACATCATGCGC 919
QY 897 CAGAAAGGATGTATGTGGCTGTGCAAGAGAGGCTGTGGAGAACCTTACTACTGTGT 956
Db 920 CAGAAAGGATGTATGTGGCTGTGCAAGAGAGGCTGTGGAGAACCTTACTACTGTGT 979
QY 957 TTAATCTCCAAAGCCATCTCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGAGAC 1016
Db 980 TTAATCTCCAAAGCCATCTCTCTCTCCATGAGAGAAAGCCGCTTTTGTGTGAGAC 1039
QY 1017 ATGCTGGCTGTGGCAAAATTTGCAATGAAAGAGCTGCACTAGGCTGTGTGTTGAC 1076
Db 1040 ATGCTGGCTGTGGCAAAATTTGCAATGAAAGAGCTGCACTAGGCTGTGTGTTGAC 1099
QY 1077 ATGATCTGACAGAGAGAAATGAGAGCTCAAAAGTCAAAAATCTCTGAAAAACGAGATT 1136
|||||
```

```
Db 1100 ATGATCTGACAGAGAGAAATGAGAGCTCAAAAGTCAAAAATCTCTGTAAGAGAGATT 1159
QY 1137 TGSCCTTCATCTCAGTGTGATATATCCCTCCCAAAAGAGAAACAGAGGCTTATCTT 1196
Db 1160 TGSCCTTCATCTCAGTGTGATATATCCCTCCCAAAAGAGAAACAGAGGCTTATCTT 1219
QY 1197 TGTGTCAAAAGAGAGAGTCAACCAATGTGTGAGAGCAAGATGCTGTGACAGTTGAG 1256
Db 1220 TGTGTCAAAAGAGAGAGTCAACCAATGTGTGAGAGCAAGATGCTGTGACAGTTGAG 1279
QY 1257 TACTTACCTTGGCTAA 1273
Db 1280 TACTTACCTTGGCTAA 1296

RESULT 14
US-09-831-426-4
: Sequence 4: Application US/09831426
: GENERAL INFORMATION:
: APPLICANT: Hoechst Marion Roussel
: APPLICANT: Bordon-Pallier, F.
: APPLICANT: Roehrer, C.
: TITLE OF INVENTION: Human hFIIIA gene and coded hFIIIA protein
: FILE REFERENCE: 146.1364
: CURRENT APPLICATION NUMBER: US/09/831,426
: CURRENT FILING DATE: 2001-05-08
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: Patentin Vers. 2.0
: SEQ ID NO: 4
: LENGTH: 1213
: TYPE: DNA
: ORGANISM: Human
US-09-831-426-4

Query Match 95.28; Score 1211.4; DB 32; Length 1213;
Best Local Similarity 99.98; Pred. No. 2,1e-265;
Matches 1212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 61 GTGCGGCGTGTGCGCGAAGGTTTCAGCAGGAGCGCGTGGCGCGCGCGGTTCCCGGC 120
Db 1 GTGCGGCGGCGCGCGGAAAGTTTCAGCAGGAGCGCGTGGCGCGCGCGGTTCCCGGC 60
QY 121 AGCTGTCTGCGGACGTTGGCAGCGCGCGCTGGCGCTTGGAGCGCGCGCGCTTGA 180
Db 61 AGCTGTCTGCGGACGTTGGCAGCGCGCGCTGGCGCTTGGAGCGCGCGCGCTTGA 120
QY 181 TCCGCGCGCGGCGTGGTGGCGGAGTGGTGTCTCTTGACCATTCGCGAGCGCTTCAATG 240
Db 121 TCCGCGCGCGGCGTGGTGGCGGAGTGGTGTCTCTTGACCATTCGCGAGCGCTTCAATG 180
QY 241 AGCGCGGAGAGCTTACGTCGCAACCGCGCGCGCGCGCTTCCAGAGAGTTTATCTG 300
Db 181 AGCGCGGAGAGCTTACGTCGCAACCGCGCGCGCGCGCTTCCAGAGAGTTTATCTG 240
QY 301 CTCCTTCCCTGACTGCAAGCGCAATTAACCAAGAGCTTGAAGCTTGAAGCGCACCTGTG 360
Db 241 CTCCTTCCCTGACTGCAAGCGCAATTAACCAAGAGCTTGAAGCTTGAAGCGCACCTGTG 300
QY 361 CAAGCAACAGGGGGAGAGACCATTTTGTGTGACTATGAAAGGTTGGCAAGCGCTTAT 420
Db 301 CAAGCAACAGGGGGAGAGACCATTTTGTGTGACTATGAAAGGTTGGCAAGCGCTTAT 360
QY 421 CAGGAGCTACCATCTGAGCGCGCACTTGTGACTGCAACAGAGAGAAAGCGCTTGTG 480
Db 361 CAGGAGCTACCATCTGAGCGCGCACTTGTGACTGCAACAGAGAGAAAGCGCTTGTG 420
QY 481 TGCAGCACTGGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAACATTTTGA 540
Db 421 TGCAGCACTGGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAACATTTTGA 480
QY 541 ACGCAACATGAATAACAAACAAATATATATGACAGTTTGAAGAGCTTGAAGAGAC 600
Db 481 ACGCAACATGAATAACAAACAAATATATATGACAGTTTGAAGAGCTTGAAGAGAC 540
```

```
QY 601 CTTTAAAGAACATCAGCAGCTGAAATCCATCAGTCCAGCATACCAATGAACCTCTATT 660
DB 541 CTTTAAAGAACATCAGCAGCTGAAATCCATCAGTCCAGCATACCAATGAACCTCTATT 600
QY 661 CAAGTGTACCCAGGAAGATGTGGGAAACACTTTTGCATCACCAGCAAGCTGAAACGACA 720
DB 601 CAAGTGTACCCAGGAAGATGTGGGAAACACTTTTGCATCACCAGCAAGCTGAAACGACA 660
QY 721 TGCCAGGCCACAGAGGCTATGTATGTCAAAAGGATGTCTTGTGGCAAAACATG 780
DB 661 TGCCAGGCCACAGAGGCTATGTATGTCAAAAGGATGTCTTGTGGCAAAACATG 720
QY 781 GACGGAACTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATGTGAAGTATG 840
DB 721 GACGGAACTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATGTGAAGTATG 780
QY 841 CGGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAACCTCATGCCACAGA 900
DB 781 CGGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAACCTCATGCCACAGA 840
QY 901 AAGGGATGTATGTCCTGTCCTCAAGAGAGGCTGTGGAAAGCTTATACTACTGTGTTTAA 960
DB 841 AAGGGATGTATGTCCTGTCCTCAAGAGAGGCTGTGGAAAGCTTATACTACTGTGTTTAA 900
QY 961 TCTCCAAAGCCATATCTCTCTCCATCAGGAGAAAGCCGCTTTGTGTGAACATGC 1020
DB 901 TCTCCAAAGCCATATCTCTCTCCATCAGGAGAAAGCCGCTTTGTGTGAACATGC 960
QY 1021 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTCAGTAGGCATGTCTGTGTACATGA 1080
DB 961 TGGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTCAGTAGGCATGTCTGTGTACATGA 1020
QY 1081 TCCTGACAAAGAAATGAAGTCAAGTCAAAATCTCGTGAAGAAACAGGCTTTGGC 1140
DB 1021 TCCTGACAAAGAAATGAAGTCAAGTCAAAATCTCGTGAAGAAACAGGCTTTGGC 1080
QY 1141 CTCTCATCTCAGTGGATATATCCCTCCCAAGGAAACAGGCAAGGCTTATCTTTGTG 1200
DB 1081 CTCTCATCTCAGTGGATATATCCCTCCCAAGGAAACAGGCAAGGCTTATCTTTGTG 1140
QY 1201 TCAAAACGGAGAGTACCCCAACTGTGTGGAGAACAGATGCTCTCGACAGTTGCAGTACT 1260
DB 1141 TCAAAACGGAGAGTACCCCAACTGTGTGGAGAACAGATGCTCTCGACAGTTGCAGTACT 1200
QY 1261 TACCCCTGGCTAA 1273
DB 1201 TACCCCTGGCTAA 1213
```

RESULT 15

US-09-831-426C-4

; Sequence 4, Application US/09831426C

; GENERAL INFORMATION:

; APPLICANT: Hoechst Marion Roussel

; APPLICANT: Bordon-Pallier, F.

; APPLICANT: Rocher, C.

; TITLE OF INVENTION: Human htfIIIA gene and coded htfIIIA protein

; FILE REFERENCE: 146.1364

; CURRENT APPLICATION NUMBER: US/09/831,426C

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: Patent In Vers. 2.0

; SEQ ID NO 4

; LENGTH: 1213

; TYPE: DNA

; ORGANISM: Human

US-09-831-426C-4

Query Match

Best Local Similarity 95.2%; Score 1211.4; DB 32; Length 1213;

Matches 1212; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 61 GTGCGCGGCTGCGCGAAGGTTTACGACAGGAGCCGTGGCGCGCGCGCGCTTCCCAGC 120
DB 1 GTGCGCGCGCGCGCGAAGGTTTACGACAGGAGCCGTGGCGCGCGCGCGCTTCCCAGC 60
QY 121 ACGTGTCTCGGACAGTGGCAGCGCGCTGGCCCTGGGCTTGGAGCGCGCGCGCTGGA 180
DB 61 ACGTGTCTCGGACAGTGGCAGCGCGCTGGCCCTGGGCTTGGAGCGCGCGCGCTGGA 120
QY 181 TGCCCGCGCGCTGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 240
DB 121 TGCCCGCGCGCTGCGCGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 241 AGCGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCTTCCAGAGGTTTCACTG 300
DB 181 AGCGCGGAGAGCTCAGCTCCGACCGCGCGCGCGCGCGCTTCCAGAGGTTTCACTG 240
QY 301 CTCCTTCTCCCTCAGCTGACGCGCAATACAGCAAGGCTGGAAGCTTGAAGCTTGAAG 360
DB 241 CTCCTTCTCCCTCAGCTGACGCGCAATACAGCAAGGCTGGAAGCTTGAAGCTTGAAG 300
QY 361 CAAGCACACGGGGAGAGACCATTTTGTGTGACTATGAAGGCTGTGGCAAGGCTTTCAT 420
DB 301 CAAGCACACGGGGAGAGACCATTTTGTGTGACTATGAAGGCTGTGGCAAGGCTTTCAT 360
QY 421 CAGGACATACCATCTGAGCGCGCACATTTCTGACTCAGCAGGAGAAAGCGTTGTTG 480
DB 361 CAGGACATACCATCTGAGCGCGCACATTTCTGACTCAGCAGGAGAAAGCGTTGTTG 420
QY 481 TGCAGCCACTGGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAACATTTTGA 540
DB 421 TGCAGCCACTGGCTGTGATCAAAATTCACACAAATCAAACTTGAAGAAACATTTTGA 480
QY 541 AGCGAAACATCAATCAACAAACAAATATATATATATATATATATATATATATATAT 600
DB 481 AGCGAAACATCAATCAACAAACAAATATATATATATATATATATATATATATATAT 540
QY 601 CTTTAAAGAACATCAGCAGCTGAAATCCATCAGTCCAGCATACCAATGAACCTCTATT 660
DB 541 CTTTAAAGAACATCAGCAGCTGAAATCCATCAGTCCAGCATACCAATGAACCTCTATT 600
QY 661 CAAGTGTACCCAGGAAGATGTGGGAAACACTTTTGCATCACCAGCAAGCTGAAACGACA 720
DB 601 CAAGTGTACCCAGGAAGATGTGGGAAACACTTTTGCATCACCAGCAAGCTGAAACGACA 660
QY 721 TGCCAGGCCACAGAGGCTATGTATGTCAAAAGGATGTCTTGTGGCAAAACATG 780
DB 661 TGCCAGGCCACAGAGGCTATGTATGTCAAAAGGATGTCTTGTGGCAAAACATG 720
QY 781 GACGGAACCTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATGTGAAGTATG 840
DB 721 GACGGAACCTTCTGAAACATGTGAGAGAAACCCATAAAGAGGAAATACTATGTGAAGTATG 780
QY 841 CGGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAACCTCATGCCACAGA 900
DB 781 CGGAAACATTTAAACGCAAGATTACCTTAAGCAACACATGAACCTCATGCCACAGA 840
QY 901 AAGGGATGTATGTCCTGTCCTCAAGAGAGGCTGTGGAAAGCTTATACTACTGTGTTTAA 960
DB 841 AAGGGATGTATGTCCTGTCCTCAAGAGAGGCTGTGGAAAGCTTATACTACTGTGTTTAA 900
QY 961 TCTCCAAAGCCATATCTCTCTCCATCAGGAGAAAGCCGCTTTGTGTGAACATGC 1020
DB 901 TCTCCAAAGCCATATCTCTCTCCATCAGGAGAAAGCCGCTTTGTGTGAACATGC 960
QY 1021 TGCGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTCAGTAGGCATGTCTGTGTACATGA 1080
DB 961 TGCGCTGTGGCAAAACATTTGCAATGAACAAAGCTCTCAGTAGGCATGTCTGTGTACATGA 1020
QY 1081 TCCTGACAAAGAAATGAAGTCAAGTCAAAATCTCGTGAAGAAACAGGCTTTGGC 1140
DB 1021 TCCTGACAAAGAAATGAAGTCAAGTCAAAATCTCGTGAAGAAACAGGCTTTGGC 1080
QY 1141 CTCTCATCTCAGTGGATATATCCCTCCCAAGGAAACAGGCAAGGCTTATCTTTGTG 1200
DB 1081 CTCTCATCTCAGTGGATATATCCCTCCCAAGGAAACAGGCAAGGCTTATCTTTGTG 1140
QY 1201 TCAAAACGGAGAGTACCCCAACTGTGTGGAGAACAGATGCTCTCGACAGTTGCAGTACT 1260
DB 1141 TCAAAACGGAGAGTACCCCAACTGTGTGGAGAACAGATGCTCTCGACAGTTGCAGTACT 1200
QY 1261 TACCCCTGGCTAA 1273
DB 1201 TACCCCTGGCTAA 1213
```

Db 1081 CTCTCATCTCAGTGGATATATCCCTCCAAAGCAAGGCAAGGCTTATCTTTGTG 1140
Qy 1201 TCAAAACGGAGAGTCACCCAACTGTGTGGAGACAAAGATGCTCTCGACAGTTGCAGTACT 1260
Db 1141 TCAAAACGGAGAGTCACCCAACTGTGTGGAGACAAAGATGCTCTCGACAGTTGCAGTACT 1200
Qy 1261 TACCCTTGGCTAA 1273
Db 1201 TACCCTTGGCTAA 1213

Search completed: February 10, 2003, 14:48:16
Job time : 3133.12 secs